Investigating Chemoresistance in Relapsed/Refractory B cell non-Hodgkin Lymphoma

Alexander Newman
Student No. 110178637

Supervisors: Professor Vikki Rand, Dr Simon Bomken, Dr Chris Bacon & Professor Christine Harrison

Northern Institute for Cancer Research

A thesis submitted in part requirement for the degree of Doctor of Philosophy from the Faculty of Medical Sciences

February 2020

Blank page

Abstract

Paediatric B-cell non-Hodgkin Lymphoma (B-NHL), namely Burkitt lymphoma and diffuse large B cell lymphoma, is successfully treated in the majority of patients in the UK at the cost of debilitating toxicity. For patients who undergo disease progression the prognosis is dire, with salvage rates as low as 20%. Previous studies have identified putative markers of disease progression, but none are currently used in the clinic. There is a clear need for usable markers of relapse/refractory disease at diagnosis for paediatric B-NHL with the aim to stratify patients and identify new potentially targetable genes and pathways.

Copy number analysis of 162 patients from the CCLG and published data identified genomic aberrations associated with disease progression: 17p copy number neutral loss of heterozygosity (CNN-LOH), 3q29 amplification and 17q CNN-LOH. 17p CNN-LOH was a prognostic marker with a hazard-ratio of 5.6 (95% CI 2-16, p=0.001, Cox proportional hazard method). *TP53* was investigated further using a combination Sanger sequencing and whole-exome sequencing. *TP53* aberrations were present in 52/95 cases, with biallelic abnormalities conferring poorer outcomes. Biallelic *TP53* aberrations were also associated with complex chromosomal abnormalities, including a novel aberration termed 13q^{plex}. Copy number analysis of 105 endemic BL patients treated in Malawi showed that prognostic aberrations in sporadic BL are present but not prognostic in endemic BL. *TP53* aberrations were identified in endemic BL and were not associated with relapse, however biallelic cases had an inferior overall survival.

Investigating 11 diagnostic and relapse pairs demonstrated that *TP53* status drives evolution of chemo-resistant disease. BLs with *TP53* aberrations at diagnosis exhibited linear evolution, while *TP53* normal cases had early-diverging patterns of progression and acquired *TP53* aberrations at relapse. We report *TP53* as an important prognostic marker in paediatric B-NHL that confers higher risk of disease progression and may help inform treatment decisions allowing for the possibility of new treatments.

Dedications

I dedicate this thesis to my Mum, Grainne. All my life you've pushed me to be more than I thought I could be, and go further than I thought possible, and it's because of you that I'm here now at the end of this project. I wish more than anything that you could have been here to see it finished and see me graduate, but I hope you'd be proud of me and what I've been able to do. This is us fighting back against cancer.

This PhD project is the hardest challenge I have ever undertaken, and I couldn't have done it without the help of my family, my friends and my great team in the Lymphoma group.

I would like to thank my family, David, Grainne and Rory, for pushing me to be the best I can be and to work hard toward my future. I'd also like to thank you for supporting me in so many ways to get me to where I am.

I would also like to thank my wonderful girlfriend Gemma, who has been with me through everything; good and bad. You've supported me through things that neither of us ever thought would happen and have been so understanding. Part of this PhD belongs to you, as you've worked just as hard keeping me going and getting me through it. Thank you as well for getting us Ash the Borador, the best dog and writing buddy I could ask for!

I would like to thank my workplace proximity associate Alex Blain, for checking in on me when things were tough. You've always been there when I needed a coffee (or a beer) and have provided some truly sage wisdom. Starting as my lowly undergrad student and rising to be my fellow PhD student, you've helped me so much along the way. I know your journey will take you to great things.

I know I would have gone crazy if not for the Dungeons and Dragons gang keeping me sane and entertained throughout the last year or so of the PhD so thank you Dino, Kent, James and Nef. I'd also like to thank Steve for taking me out of the writing world to play games and have some fun every now and then, and Joe for always being around to go out and play some Pokémon Go. Thank you, Melanie, for being so supportive as we took on both our MRes and PhD projects together, we made it!

I would like to thank the Lymphoma Genomics Group, or ‘Team Lymphoma”. We’re a tight-knit group who always have each other’s backs, and without them I would have absolutely fallen by the wayside. Peixun, you’ve helped me with more questions in the lab than I could ever remember, and taught me everything I know! Masood, you’ve helped me become a more powerful bioinformatician than I could possibly imagine. I want to thank Simon for being a such a great supervisor, helping to push this project in the right direction and giving such helpful advice and support throughout. Thank you, Chris, for working hard to help me improve my work and giving me a better eye for detail.

I would like to thank Vikki (The Professor) for taking a chance on a young, eager MRes student who had an interest in Burkitt lymphoma and giving me the opportunity to develop into a well-balanced researcher. You’ve pushed me at the right times to make sure my work is better than I could have imagined, and you’ve been extremely patient and understanding with me during my darkest days. Thank you.

I would like to thank Ros, Ian, and all the other wonderful people at NEPAC (North East Promenaders Against Cancer) who have worked tirelessly to raise money so that we can beat cancer, and have sponsored my project. I hope that my project does all your hard work justice.

Acknowledgements

I would like to thank Professor Vikki Rand for the conception of the project and generating the initial data that allowed this project to become a reality.

Extraction of DNA and RNA for this project from patient material was performed by myself, Dr Peixun Zhou, Dr Amy Erhorn, Dr Rachel Crossland, Alex Blain and Amy Barnard.

FISH slide analysis was performed by myself, Dr Amy Erhorn, Dr Peixun Zhou and Amy Barnard.

Next generation sequencing library prep was performed by AROS (Applied Biosystems, Germany) and the Core Genomics Facility (Centre for Life, Newcastle, UK).

Whole exome sequencing analysis, from pre-processing to variant calling, was carried out by myself and Dr Masood Zaka, with help from Dr Matthew Bashton.

Circos plots presented in Chapter 4 were generated by Dr Masood Zaka.

TP53 Sanger sequencing data for the endemic BL cohort used in Chapter 6 was performed by Dr Peixun Zhou alongside my project.

Pathology review of patient material was performed by Drs Christopher Bacon, Despina Televantou and Katrina Wood. Unstained sections were cut from paraffin blocks by the Cellular Pathology department at the Royal Victoria Infirmary, Newcastle, UK.

Clinical data collection for the CCLG cohort was kindly performed by dedicated research nurses and clinicians at individual centres.

Blank page

Table of Contents

Table of Contents

List of Figures

List of Tables

Abbreviations

Abbreviation	Definition
ABC	Activated B cell
aCGH	Array comparative genome hybridisation
ALCL	Anaplastic large cell lymphoma
AML	Acute myeloid leukaemia
ATP	Adenosine triphosphate
B-ALL	B cell acute lymphoblastic leukaemia
B-HLH	Basic helix-loop-helix
B-NHL	B cell non-Hodgkin lymphoma
B-NHL, NOS	B cell non-Hodgkin lymphoma, not otherwise specified
BAC	Bacterial artificial chromosome
BAF	B allele frequency
BAM	Binary Alignment Map
BCLU	B cell lymphoma unclassified
BCP-ALL	B cell precursor acute lymphoblastic leukaemia
BL	Burkitt lymphoma
BLAST	Basic local alignment search tool
BLAT	BLAST-like alignment tool
BLL-11q	Burkitt-like lymphoma with 11q aberration
BM	Bone marrow
bp	Base pair
BWA	Burrows-Wheeler Aligner
BWA-MEM	Burrows-Wheeler Aligner, maximal exact matches
CBS	Circular binary segmentation
CCLG	Children's Cancer and Leukaemia Group
CDK	Cyclin-dependent kinase
CEP7	Centromere Probe Chromosome 7
CGH	Comparative genome hybridisation
CHOP	Cyclophosphamide, doxorubicin, vincristine, prednisolone
CLL	Chronic lymphocytic leukaemia
CNN-LOH	Copy number neutral loss of heterozygosity
CNS	Central nervous system
COG	Children's Oncology Group
COPADM	Cyclophosphamide, vincristine, prednisolone, doxorubicin, methotrexate
CSF	Cerebrospinal fluid
CSR	Class switch recombination
CYM1	Cytarabine, methotrexate
CYVE2	Cytarabine, etoposide
DAPI	4’,6-diamidino-2-phenylindole
dbGaP	Database of genotype and phenotype
DLBCL	Diffuse large B cell lymphoma
DNA	Deoxyribonucleic acid
DNE	Dominant negative effect

dNTP	Deoxyribonucleic triphosphate
EBER	Epstein-Barr virus encoded small RNAs
eBL	Endemic BL
EBV	Epstein-Barr virus
EFS	Event-free survival
ERIC	European Research Initiative on CLL
FAB/LMB-96	French-American-British Lymphoma Malignancy B, 1996
FASST2	Fast adaptive state segmentation 2
FASTQ	Sequencing read file
FF	Fresh frozen
FFPE	Formulin-fixed, paraffin-embedded
FISH	Fluorescence in situ hybridisation
FL	Follicular lymphoma
GATK	Genome analysis toolkit
GC	Germinal centre
GCB	Germinal centre B cell
GEO	Gene Expression Omnibus
GEP	Gene expression profile
GISTIC	Genomic Identification of Significant Targets in Cancer
GISTIC2.0	Genomic Identification of Significant Targets in Cancer 2.0
GPCR	G protein coupled receptor
HC	Haplotype Caller
HD	High definition
HE	Haematoxylin and eosin
HGBL, NOS	High grade B cell lymphoma, not otherwise specified
HIV	Human Immunodeficiency Virus
HLA	Human leukocyte antigen
HMM	Hidden Markov Model
HMRN	Haematological Malignancy Research Network
HR	Hazard ratio
HTA	Human Tissue Act
iAMP21	Interstitial amplification of chromosome 21
IARC	International Agency for Research in Cancer
IGH	Immunoglobulin heavy chain
IGK	Immunoglobulin light chain kappa
IGL	Immunoglobulin light chain lambda
IGV	Integrative Genomics Viewer
IHC	Immunohistochemistry
IPI	Independent Prognostic Index
L3ALL	Burkitt's acute lymphocytic leukaemia
LFL	Li-Fraumeni-like
LFS	Li-Fraumeni Syndrome
LR	Likelihood ratio
MCR	Minimal common region
MgCl2	Magnesium chloride

MIP	Molecular inversion probe
miRNA	Micro RNA
MMP	Matrix metalloproteases
mRNA	Messenger RNA
NCI	National Cancer Institute
NEPAC	North East Promenaders Against Cancer
NFkB	Nuclear factor Kappa B
NLOD	Normal Log Odds
NOS	Not otherwise specified
NP-40	nonyl phenoxypolyethoxylethanol
OS	Overall Survival
PCR	Polymerase chain reaction
PhD	Doctorate of Philosophy
PI3K	Phosphoinositide 3-kinases
PON	Panel of Normals
PRC2	Polycomb repressive complex 2
QC	Quality control
QECH	Queen Elizabeth Central Hospital
R-CHOP	Rituximab, cyclophosphamide, doxorubicin, vincristine, prednisolone
R-CODOX-M-IVAC	Cyclophosphamide, doxorubicin, vincristine, methotrexate/ifosfamide, etoposide, high dose cytarabine
REAL	Revised European-American Lymphoma
RNA	Ribonucleic acid
RNAi	RNA interference
RR	Risk of Relapse
RVI	Royal Victory Infirmary
SAM	Sequence Alignment Map
sBL	Sporadic BL
SGE	Son of Grid Engine
SHM	Somatic hypermutation
shRNA	Short hairpin RNA
SNP	Single nucleotide polymorphism
SNP-FASST2	Single nucleotide polymorphism fast adaptive state segmentation 2
SSC	Saline-sodium citrate
TAE	Tris-acetate
TLOD	Tumour Log Odds
TTP	Time to Progression
TYA	Teenagers and young adults
UK	United Kingdom
USA	United States of America
VAF	Variant allele frequency
WHO	World Health Organisation

Chapter 1: Introduction

B cell non-Hodgkin Lymphoma (B-NHL) is a solid haematological malignancy of B lymphocytes that presents in lymphoid tissue and other extranodal sites. Currently 34 types of mature B cell neoplasm are recognised by the World Health Organisation (WHO) (Swerdlow *et al.*, 2016) and these vary widely in incidence, median age at diagnosis, clinical presentation and prognosis. The fastest growing and most aggressive forms of mature B-NHL are Burkitt Lymphoma (BL) and diffuse large B cell lymphoma (DLBCL), while follicular (FL) and marginal zone lymphomas, for example, are considered indolent, or less aggressive.

The expected incidence of haematological neoplasms in the UK is 38,090 cases per year, of which 9620 are B-NHL based on data for 2001 recorded by the Haematological Malignancy Research Network Census (HMRN, 2019). Slightly more than half of B-NHL cases in the UK are diagnosed as DLBCL (4980 cases). The relative incidences of new B-NHL diagnoses are not consistent in all age groups (Figure 1.1). In younger patients, BL is the most common diagnosis of B-NHL until the adolescent years, where the incidence of DLBCL increases and remains the most frequently diagnosed subtype in adult patients.

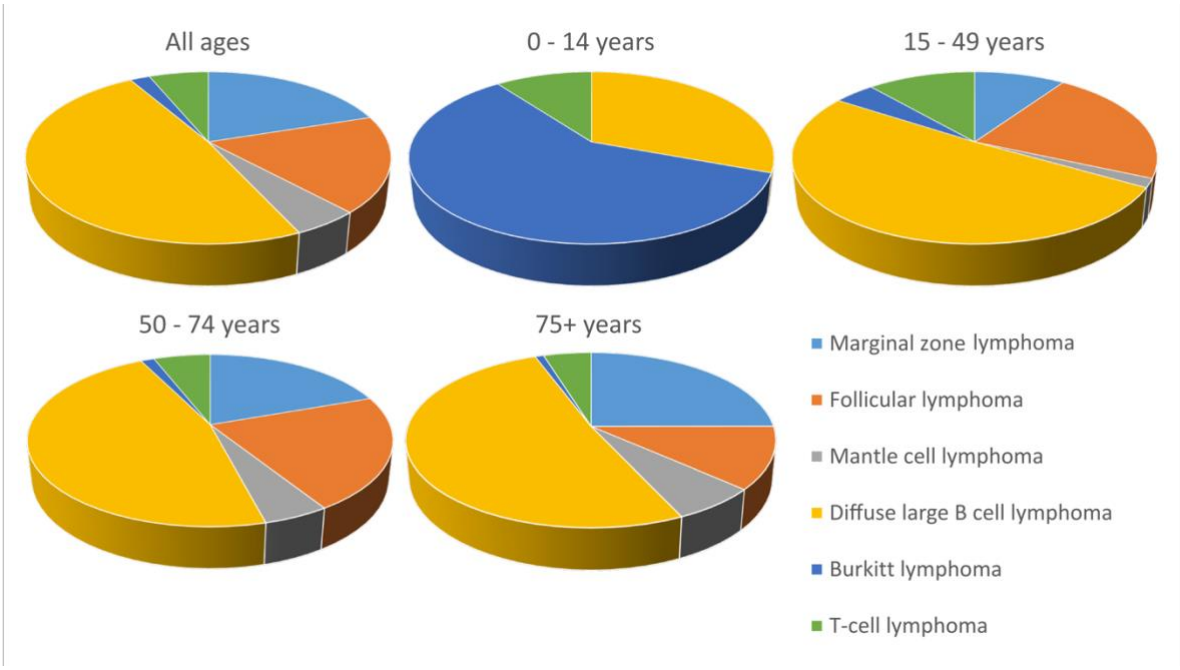


Figure 1.1 - Number of UK cases of NHL subtypes per year by age. Data from HMRN, 2019.

1.1 The Germinal Centre

The germinal centre (GC) is a transient lymphoid histological structure formed as a result of T cell-dependent antigen presentation, and many of the mechanisms driving lymphomagenesis play a role in normal GC biology (Basso and Dalla-Favera, 2015). The major role of the GC is to produce and select B cells with high affinity antibodies through somatic hypermutation. The GC is comprised of the "dark zone" and the "light zone". B cells ($CXCR4^{hi}CD83^{low}CD86^{low}$) in the dark zone are highly proliferative and undergo somatic hypermutation here (Victora *et al.*, 2010). Conversely, light zone B cells ($CXCR4^{low}CD83^{hi}CD86^{hi}$) are accompanied by dendritic cells, macrophages and T cells, and are activated and selected for based on their receptor affinity (Figure 1.2). *MYC* is an important regulator of the GC physiology, both in lymphomagenesis and normal function. *MYC* is highly expressed early in the GC initiation stages but it is then suppressed in dark zone B cells by *BCL6*. Light zone B cells that return to the dark zone by cyclic re-entry are known to re-express *MYC* (Basso and Dalla-Favera, 2015).

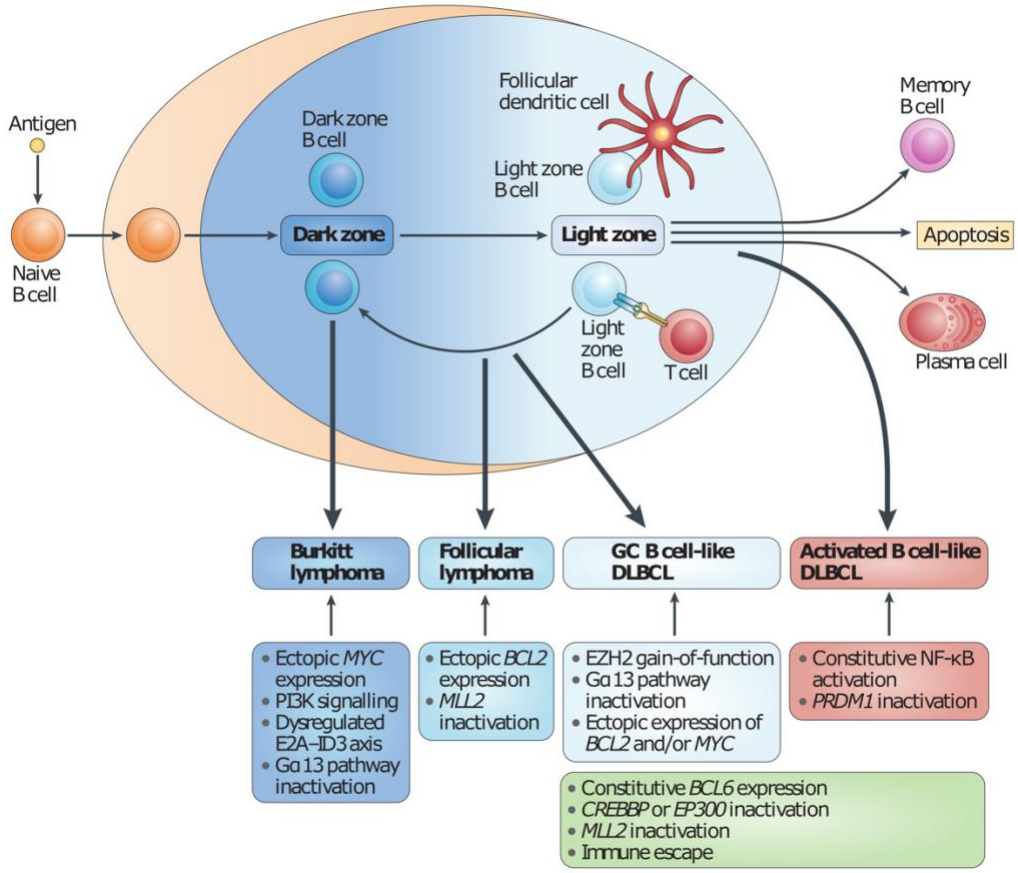


Figure 1.2 – A summary of the activity and cellular composition of the germinal centre. Taken from Basso and Dalla Favera (2015).

BCL6 is a major regulator of the GC, inhibiting target gene expression via DNA binding and recruitment of histone deacetylase complexes (Hatzi *et al.*, 2013). The *BCL6* gene is responsible for modulation of a large and complex network of genes associated with multiple related pathways including DNA damage response, somatic hypermutation, cell cycle arrest and apoptosis (Basso and Dalla-Favera, 2015). *BCL6* is crucial to adaptive immune response mechanisms and BCL6-deficient mice have been shown to lead to a complete absence of GC formation (Dent *et al.*, 1997). E2A, encoded by the *TCF3,* gene is expressed in the dark zone, driving cell cycle progression via cyclin D3 (*CCND3*) and *E2F2* induction and downregulation of *RB1*. *ID3* functions to inhibit active E2A, forming a vital axis of regulation. The interaction between *TCF3*, *ID3* and *CCND3* has been shown to be important in Burkitt Lymphoma, where 74% of cases were shown in one study to have a mutation of either *ID3* or *TCF3* (Lopez et al., 2019).

B cells undergo somatic hypermutation in the dark zone of the GC, then migrate to the light zone for antigen selection. Exit of the light zone relies on the downregulation of *BCL6* in B cells driven by BLIMP1 (*PRDM1*) expression.

The *MYC* proto-oncogene is implicated in the initiation and maintenance of many cancer types (Dang, 2012). In fact, MYC was the first gene to be shown to be associated in GC-derived lymphomagenesis (Dalla-Favera *et al.*, 1982). Expression of MYC is required for initiation of the GC and then is repressed in dark zone cells as a result of BCL6 expression (Calado *et al.*, 2012). MYC re-expression is then only observed in B cells re-entering the dark zone from the light zone.
Translocations involving immunoglobulin genes are frequently observed in mature B cell lymphomas, and this is linked to normal processes involved in B cell development including class switch recombination, V(D)J recombination and somatic hypermutation. In V(D)J recombination, induction of the recombinant activation genes *RAG1* and *RAG2* leads to cleavage of the DNA between recombination signal sequences (RSS). This mechanism in healthy cells drives antibody antigen specificity. However this machinery can lead to the juxtaposition of oncogenes adjacent to an immunoglobulin locus, usually *IGH*. The frequent involvement of IGH and other IG loci is due to its propensity to harbour double strand breaks. One of the

earliest detectable genetic events in B cell lymphoma is the t(14;18) translocation involving the *BCL2* and *IGH* genes often seen in follicular lymphoma and less frequently in GCB-DLBCL (Blombery *et al.*, 2015). This is thought to take place during the pro-B cell stage as a result of D-J rearrangement. Further translocations identified in BL and DLBCL arise within the germinal centre. The hallmark BL translocation t(8;14) involving the *MYC* and *IGH* genes occur in the dark zone as a result of aberrant class switch recombination at a time when MYC expression is still high (Dominguez-Sola *et al.*, 2012). Further evidence implicating class switch recombination is the fact that the breakpoints of in the *MYC* gene are clustered at the AID (activation induced cytidine deaminase) hotspots adjacent to exon 1 of the gene. Further alterations to the *MYC* gene are regularly observed as a result of AID-induced somatic hypermutation (Dominguez-Sola *et al.*, 2012). With the exception of high *MYC* expression in the germinal centre driven by *MYC* rearrangement, BL cells share a gene expression profile with normal dark zone germinal B cells. This pattern of expression is implemented by activating mutations in *TCF3* and inactivating mutations in *ID3*.

Ectopic *MYC* expression drives proliferation and progression through the cell cycle which supports lymphomagenesis, but also activates pro-apoptotic pathways, including via *TP53*. Genetic aberrations involving the *TP53* gene or its regulators are required to inhibit these pro-apoptotic signals and maintain lymphoma development (O'Leary *et al.*, 2004). Overexpression of *MDM2* has been reported in BL (Capoulade *et al.*, 1998), and *MDM4* expression is believed to be high in almost all BL cases (Leventaki *et al.*, 2012).

1.2 Subtypes of B-NHL

This thesis focuses primarily on two common subtypes of B-NHL: BL and DLBCL. Both present similarly but pathologists are able to differentiate between the two primarily though immunohistochemistry (IHC) and fluorescence *in-situ* hybridisation (FISH) for *MYC*-rearrangement with an immunoglobulin locus. Despite this there is still ambiguity in some B-NHL diagnoses where a firm subtype cannot be determined. In addition, intermediate diagnoses do exist, which are not fully characterised currently. For example, many paediatric B-NHL express markers that resemble BL

when queried by IHC, but lack a detectable *MYC*-rearrangement (Leucci *et al.*, 2008). Some of these cases are determined to be BL, some would have been considered “Burkitt-like” or B cell lymphoma, unclassified (BCLU) in previous WHO classifications but would now be classified as either the high grade B cell lymphoma, not otherwise specified (HGBL, NOS) or Burkitt-like lymphoma with 11q aberration (BLL-11q) which are both new to the WHO 2016 classification (Swerdlow *et al.*, 2016) (Table 1.1).

B-NHL Subtype	Abbreviation	Genetic Lesions
Burkitt lymphoma	BL	Hallmark IG-*MYC* translocation. Activating *CCND3* and *TCF3* mutations. Inactivating *ID*3 mutations
Germinal centre DLBCL	GCB-DLBCL	Frequent IGH-BCL2 translocations. BCL6 deregulation through mutations in *BCL*6, *FBXO1*1, *MEF2*B, *CREBB*P and *EP30*0. Activating *EZ*H2 mutations. MYC upregulation through *P*TEN loss. Dysregulated inhibition of centroblast migration and Akt signalling via inactivating mutations *of RHOA, GNA13* a*nd S1*PR2.
Activated B cell DLBCL	ABC-DLBCL	NF-kB activation via activating *MYD88* mutations and biallelic loss of *TNFAIP3.* BCL6 deregulation via IGH-BCL6 translocations and translocations with non-IGH partners. Failure of post-GC differentiation through *PRD*M1 mutation or loss.
Burkitt-like lymphoma with 11q aberration	BLL-11q	Lack of detectable IG-MYC translocation. Copy number aberration involving telomere of 11q.
High grade B cell lymphoma, not otherwise specified	HGBL, NOS	Lack of detectable IG-*MYC*, IG-*BCL2* and IG-*BCL6* translocations

Table 1.1 – A table summarising the main B-NHL subtypes and their associated genetics. Information collated from Blombery et al. (2015) and Swerdlow et al. (2016).

1.2.1 Burkitt lymphoma

There are three subtypes of Burkitt Lymphoma recognised by the WHO: endemic BL (eBL), sporadic BL (sBL) and immunodeficiency-associated BL (Swerdlow *et al.*, 2016). Endemic BL is primarily found in sub-Saharan Africa, accounting for approximately 50% of all childhood cancers in Africa (Molyneux *et al.*, 2012; O'Callaghan-Gordo *et al.*, 2016). Sporadic BL is the subtype found primarily in the western world, including Western Europe and the United States of America, but is diagnosed globally. The third form is associated with immunodeficiency and is recognised and specific to individuals with compromised immune systems due to transplant related immunosuppressant therapy, HIV virus infection, congenital immunodeficiency disorders or various iatrogenic causes. The hallmark of Burkitt Lymphoma, regardless of subtype, is a translocation that places the *MYC* proto-oncogene under the regulation of an immunoglobulin chain locus (most commonly *IGH*) leading to constitutive, high level expression, but less than 5% have no detectable rearrangement of the *MYC* gene (Swerdlow *et al.*, 2016). Burkitt Lymphoma is the most common aggressive lymphoma in children, but incidence decreases with age (Figure 1.1). Paediatric BL is treated successfully in a paediatric setting with over 90% 5 year event free survival (Giulino-Roth and Goldman, 2016; Minard-Colin *et al.*, 2016). Conversely, adult Burkitt Lymphoma has been reported to have 2 year event free survivals as low as 37% (Perkins and Friedberg, 2008).

1.2.2 Diffuse large B cell lymphoma

Diffuse large B cell lymphoma makes up the largest proportion of B-NHL overall and encompasses neoplasms with a wide spectrum of clinical and pathological features. DLBCL usually presents as a rapidly growing mass at a lymph node or extra-nodal site. The vast majority of DLBCL diagnoses are made in adults but the disease is found in children and adolescents. As a large and heterogeneous group of lymphomas it has been difficult historically to find subgroups to stratify patients as no method of subdivision, including morphology, was found to be reliable or reproducible by the REAL (Revised European-American Lymphoma) classification. The WHO recognises many subtypes of DLBCL including ALK+, EBV+, plasmablastic and IRF4 rearranged while cases that do not fit into an existing category are deemed to be DLBCL, NOS (not otherwise specified). More recently, gene expression profiling by microarray analysis has been utilised to identify three distinct DLBCL, NOS subtypes with expression profiles that correspond to germinal centre B cells (GCB), activated B cells (ABC) and type 3/unclassifiable (Alizadeh *et al.*, 2000), with ABC-DLBCL being

associated with a significantly poorer prognosis. The 2016 WHO classification takes elements of these previous classifications into account (Swerdlow *et al.*, 2016) and is generally adopted by clinicians and researchers across the world. GCB-like DLBCL is the most common DLBCL subtype and has a generally superior prognosis (Alizadeh *et al.*, 2000). Conversely, ABC-like DLBCL is less common and has a worse outcome. ABC-like DLBCL is also significantly less frequent in the paediatric disease compared to GCB-like. In addition to ABC-like gene expression pattern, expression of MYC (Chang *et al.*, 2000), BCL2 (de Leval and Harris, 2003) and Ki67 (Miller *et al.*, 1994) are all associated with inferior outcome. However, in paediatrics these markers are not associated with prognosis (Miles *et al.*, 2008). The most recently proposed subgroup is large B cell lymphoma with IRF4 rearrangement. The subtype is characterised by a high expression pattern of IRF4/MUM1, usually but not exclusively with an IRF4 rearrangement. It is an uncommon subgroup, present in approximately 0.05% of DLBCL diagnoses and has been observed in both adult and paediatric disease (Swerdlow *et al.*, 2016).

1.2.3 Burkitt-like Lymphoma with 11q aberration

The WHO Classification for 2016 introduced a new subtype of B-NHL characterised by a lack of IG/MYC translocation, Burkitt-like morphology and a chromosomal abnormality on 11q (BLL-11q) (Swerdlow *et al.*, 2016). Originally characterised by Salaverria *et al.* (2014) the subtype describes a subset of cases which would previously have been described as either a MYC-negative BL or BCLU (B cell lymphoma, unclassified) with Burkitt-like features. They are typically CD20+, CD10+, BCL2- and Ki67+ near 100%. The hallmark abnormality on 11q involves a gain most commonly between 11q13 and 11q23 followed by a loss of 11qter and the abnormality is most frequently described as a tandem duplication. The subtype strongly resembles a form of post-transplant Burkitt Lymphoma with no detectable *MYC* rearrangement and proximal gain of 11q paired with telomeric deletion (Ferreiro *et al.*, 2015). It is important to note that the inclusion of the BLL-11q subtype in the 2016 WHO Classification is provisional as there is still controversy regarding the true nature of the subtype, with some believing it is a BL-like subtype while others consider it a form of DLBCL or HGBL (Ferreiro *et al.*, 2015; Havelange *et al.*, 2016a). This highlights the need for better characterisation of this rare subgroup of B-NHL so that the most appropriate therapy can be given to patients and new targeted

therapies can be implemented based on robust biological understanding of these cases. Relatively little is known about the prognosis of BLL-11q as few cases have been characterised to date, however in the adult setting it is believed that BL-like cases with no detectable *MYC* rearrangement have a better prognosis on BL regiments rather than R-CHOP in the adult setting (Sevilla *et al.*, 2007). Cases of this subtype are usually EBV negative and have relatively quiet genomes, comparable to *MYC*-rearranged BL (Gonzalez-Farre *et al.*, 2019). BLL-11q cases have been reported to have frequent 5q21.3-5q32 copy number gain and deletions on 6q, but lack the 1q gains frequently seen in *MYC*-rearranged BL. Mutations in typical driver genes in BL, namely *ID3*, *TCF3* and *CCND3*, were not identified in BLL-11q cases, instead a GCB-DLBCL-like mutational landscape was observed, with mutations of *BTG2*, *ETS1*, *EP300* and *GNA13* identified more frequently. Further analysis in the same study showed that the mutational profile of BL patients without MYC rearrangement or 11q aberration resembled neither MYC-rearranged BL nor BLL-11q (Gonzalez-Farre *et al.*, 2019).

1.3 Genomic and Molecular Landscape of B-NHL

1.3.1 Burkitt Lymphoma

Burkitt Lymphoma is defined by a hallmark *MYC* translocation, predominantly t(8;14)(q24;q32) involving the immunoglobulin heavy chain (*IGH*) locus. In a smaller percentage of cases *MYC* translocations involve the immunoglobulin light chain loci lambda (*IGL*) by a t(8;22)(q24;q11.2) translocation or kappa (*IGK*) by t(2;8)(p12;q24) rearrangement to the same effect. Other than this primary translocation the BL genome is generally quiet, without additional translocations and few other major cytogenetic abnormalities (Hummel *et al.*, 2006). Burkitt Lymphoma cells resemble normal germinal centre B cells in terms of gene expression profile (Dave *et al.*, 2006). By a combination of sequencing and staining techniques, it has been shown that BL cells are most comparable to dark zone centroblastic cells, while the majority of other B cell lymphomas resemble light zone centrocytes. Very few differences in gene expression have been found between these two cell types. A notable exception is the finding that *TCF3* is upregulated in dark zone cells compared to light zone, and mutations of the gene are implicated in BL pathogenesis (Schmitz *et al.*, 2012).

IGH/MYC translocation in sporadic BL is believed to be a result of aberrant class switch recombination (CSR), the process by which the B cell uses double stranded breaks in the IGH region to change the arrangement of heavy chain genes to switch the immunoglobulin class that the cell produces (Burmeister *et al.*, 2013). Both BL and normal GC B cells express *AID*, an enzyme that deaminates cytosine bases converting them to uracil. This leads to a base mismatch, which will be bound by proteins of the mismatch repair (MMR) pathway, which in turn recruit an exonuclease to excise the aberrant uracil base and create a single strand break. Binding sites for *AID* are abundant in the IGH locus, leading to frequent deamination of cytosine bases (Ramiro *et al.* 2004). Double stranded breaks occur when two deaminated bases are close to one another on different strands. This is utilised by the B cell to excise heavy chain genes that code for now-redundant heavy chain classes (Kumar *et al.*, 2014). However it is well known now that *AID* has many off-target effects in loci outside the IGH region, including *PAX5*, *BCL6* and *MYC* (Kato *et al.*, 2012) which have been shown to lead to translocations (Lieber *et al*. 2016).

MYC acts as a transcription factor for many genes, many of which are involved in cell survival, proliferation and apoptosis, resulting in its implication in a plethora of human cancers. It was first identified as a viral gene in the genome of the acute transforming retrovirus which caused tumours in chickens (Duesberg and Vogt, 1979). The identification of *MYC* at the breakpoint of t(8;14) translocations in BL led to the confirmation of its role as a proto-oncogene. Transgenic mouse experiments show that *MYC* is transforming *in vivo*, but that cessation of *MYC* expression leads to regression of tumours, suggesting that continuous expression is required to maintain the tumour state, and that inactivation of *MYC* could be a viable therapeutic target. Additionally, due to the clonal nature of transgenic models with ectopic expression of *MYC* it is suggested that secondary abnormalities are required in BL for malignant transformation to occur. Additionally, transgenic Eμ-MYC mouse models of lymphoma are distinctly different to BL, often being lymphoblastic (Harris *et al.*, 1988). Eμ-MYC mouse models have been used extensively to investigate the pathogenesis of *MYC*-driven lymphomagenesis. A study using these mouse models showed that constitutive activation of *MYC* accompanied with deregulated PI3K activity gives rise to BL-like tumours (Sander *et al.*, 2012).

Several studies have investigated the genomic landscape of Burkitt Lymphoma using a range of low- and high-resolution techniques. The vast majority of studies have primarily described sporadic BL. Poirel *et al.* (Poirel *et al.*, 2009) collected karyotype data from 238 BL and DLBCL patients treated on the FAB/LMB 96 trial (French, American British/Lymphomes Malins B). The most frequently identified abnormality in BL, other than *MYC* translocation was gain of 1q (29% of patients), followed by gain of 7q (14% of patients) and deletion of 13q (14% of patients). Copy number gains involving chromosome 1q have been reported in several studies as the most frequent secondary abnormality in BL (Lones *et al.*, 2004; Onciu *et al.*, 2006; Toujani *et al.*, 2009; Scholtysik *et al.*, 2010) with conflicting conclusions regarding prognostic value. Poirel *et al.* had the largest patient cohort but did not detect any significant difference in prognosis between 1q gained and 1q normal BL patients. However, sample size and varying therapeutic regimens may go some way to explain the differences in outcome.

Chromosome 13 is frequently aberrant in B cell lymphoma, including BL. Cytogenetic studies have identified deletions of 13qter in BL. Lones *et al.* (2004) detected loss of 13q32-qter in two patients (2/18, 11%) and both of these patients relapsed while the abnormality was not detected in any of the patients who did not relapse. Poirel *et al.* (2009) detected 13q loss in 38/238 (16%) cases, with the most commonly deleted region mapped to 13q34. Remarkably, 13q loss did not differ significantly in incidence between BL and DLBCL cases, suggesting involvement of a pathway common to both disease subtypes. An investigation utilising FISH (fluorescence *in situ* hybridiation) to characterise chromosome 13 and 7 abnormalities in Burkitt Lymphoma and evaluate their prognostic significance identified 13q deletions in 42% of patients, considerably more frequently than the Poirel study (Poirel *et al.*, 2009; Nelson *et al.*, 2010). The study observed 20% of cases (18/90) with deletion of both 13q14.3 and 13q34 and an additional 8% harboured a loss of 13q34 alone (7/90). Karyotyping data was available for 18 of the cases investigated, and 44% of these (8/18) showed loss of 13q compared to 78% by FISH of the same 18 samples. This partial lack of consistency between techniques may explain the difference in frequency of 13q loss between studies. 13q14.3 deletion was associated with significantly inferior prognosis and it was hypothesised that deletion of *DLEU1* and *DLEU2* conferred the poor outcome in these patients (Lee *et al.*, 2017).

The increasing resolution of microarray technology enabled greater understanding of 13q abnormalities in B cell lymphoma. Array CGH analysis of 15 BL cell lines and 13 patient samples (Toujani *et al.*, 2009) identified a small minimal common region (MCR) of amplification at 13q31.3-q32.1. The authors identified this amplified region in five cell lines and two tumour samples. The region contained 12 genes and one microRNA cluster, and it was postulated that *ABCC4* was the driver gene in this region. *ABCC4* is a member of the ATP-binding cassette gene superfamily of transporter proteins and is part of the MRP (multi-drug resistance-associated protein) subfamily, which is known to be involved in multiple drug resistance in cancer that is reversed when the gene is silenced by RNAi (Zhang *et al.*, 2015). However later genomic studies with higher resolution SNP arrays refined the region of amplification of the 13q31.3 region of amplification in BL and postulated that it was the amplification of miRNA cluster *MIR17HG* that was most significant within this region (Scholtysik *et al.*, 2010; Schiffman *et al.*, 2011). Schiffman *et al*. (Schiffman *et al.*, 2011) reported that of 28 paediatric Burkitt Lymphoma samples analysed by SNP array, three harboured amplification of *MIR17HG* and of them two relapsed (Figure 1.3). This raised the possibility of a prognostic role for *MIR17HG* amplification in paediatric BL but this requires further investigation in a larger paediatric cohort.

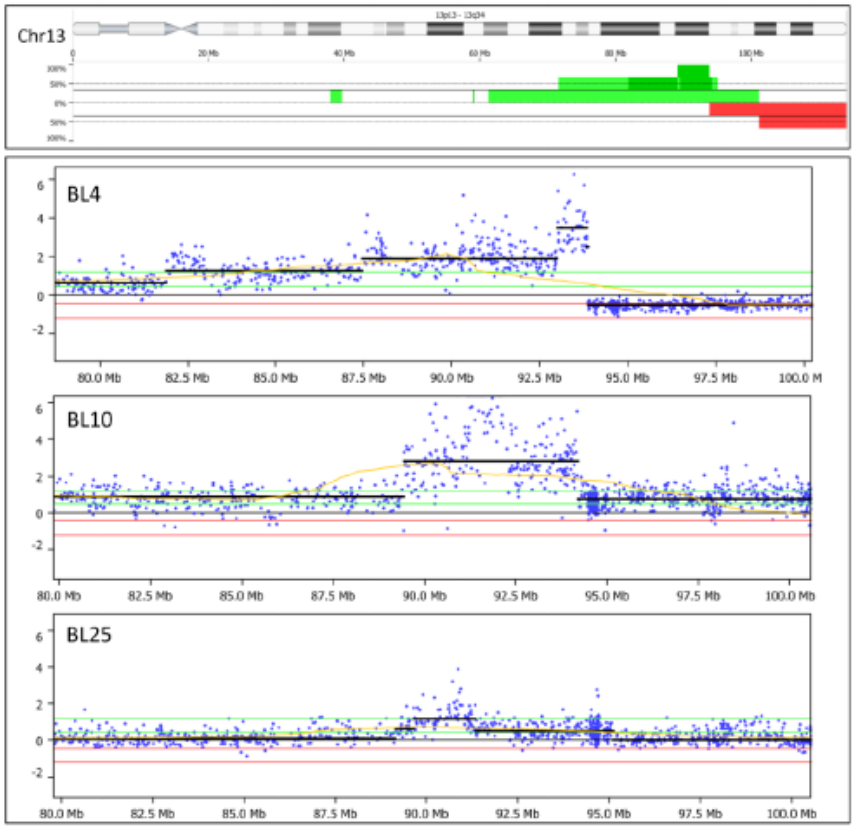


Figure 1.3 - Copy Number Gain of MIR17HG in three paediatric BL cases. Adapted from Schiffman et al*. 2011.The top panel shows the size and nature of copy number gains and deletions on chromosome 13q. Green bars represent copy number gain and red bars represent deletion. Panels 2-4 show the processed 13q copy number data for the three patients.*

A recent study further investigated the amplification of *MIR17HG* in BL using Affymetrix Mapping 100K SNP arrays and Fluorescence *in Situ* Hybridisation (FISH). A cohort of 24 patients (13 children and 11 adults) was screened for the abnormality (Havelange *et al.*, 2016b). *MIR17HG* amplification was detected in five patients (20.8%). Interestingly, four of the five patients were children and the one adult was 22 years old, suggesting that the abnormality may be an age-associated abnormality in Burkitt Lymphoma. However, they did not include enough patients in the study to meaningfully investigate the prognostic impact of the abnormality. Other than the incidence of *MIR17HG* amplification no significant differences in copy number profiles between adult and paediatric BL have been reported (Klapper *et al.*, 2012).

As described earlier, gain of chromosome 7 is a frequent abnormality in BL and has been shown to have a significant negative effect on event-free survival (83.6% vs 72.2%) (Poirel *et al.*, 2009). The region contains 64 genes and, notably, the *mir106b-25* miRNA cluster, a paralog of the *MIR17HG* miRNA cluster. Like *MIR17HG*,

MIR106B is involved in proliferation and resistance to apoptosis and has been shown to induce doxorubicin resistance in breast cancer cells. Nelson *et al*. (2010) investigated the incidence and prognostic value of chromosome 7 gains in 68 BL in patients from the COG (Children's Oncology Group) CCG-5961 study. 10% of patients (7/68) investigated by FISH for CEP7 (centromere of chromosome 7) harboured a gain of the region. This result is consistent with the findings of Poirel *et al.* despite investigating a different region of the chromosome. Results from the Poirel study suggested that smaller regions of copy number gain were detected proximal to 7q21-q22 which would not be identified by a CEP7 probe. Nelson *et al.* did not detect a significant difference in overall survival or event-free survival in patients with chromosome 7 gains compared to the rest of the cohort, conflicting with previously published studies (Lones *et al.*, 2004; Poirel *et al.*, 2009). A possible explanation for this is that both studies treated patients using different therapy protocols: the Poirel study incorporated patients treated as part of the FAB/LMB96 trial with COPADM, CYVE/CYM while the Nelson paper studied patients treated with the much less intensive CHOP regimen. This would not explain the conflicting incidences of abnormalities between studies but may explain difference in prognostic value.

Four studies investigating the mutational spectrum of Burkitt Lymphoma were published in 2012 using next generation sequencing approaches (Love *et al.*, 2012; Richter *et al.*, 2012; Schmitz *et al.*, 2012; Giulino-Roth and Goldman, 2016). The results of these studies, as well as two subsequent sequencing studies (Abate *et al.*, 2015; Havelange *et al.*, 2016b) are summarised in Figure 1.4.

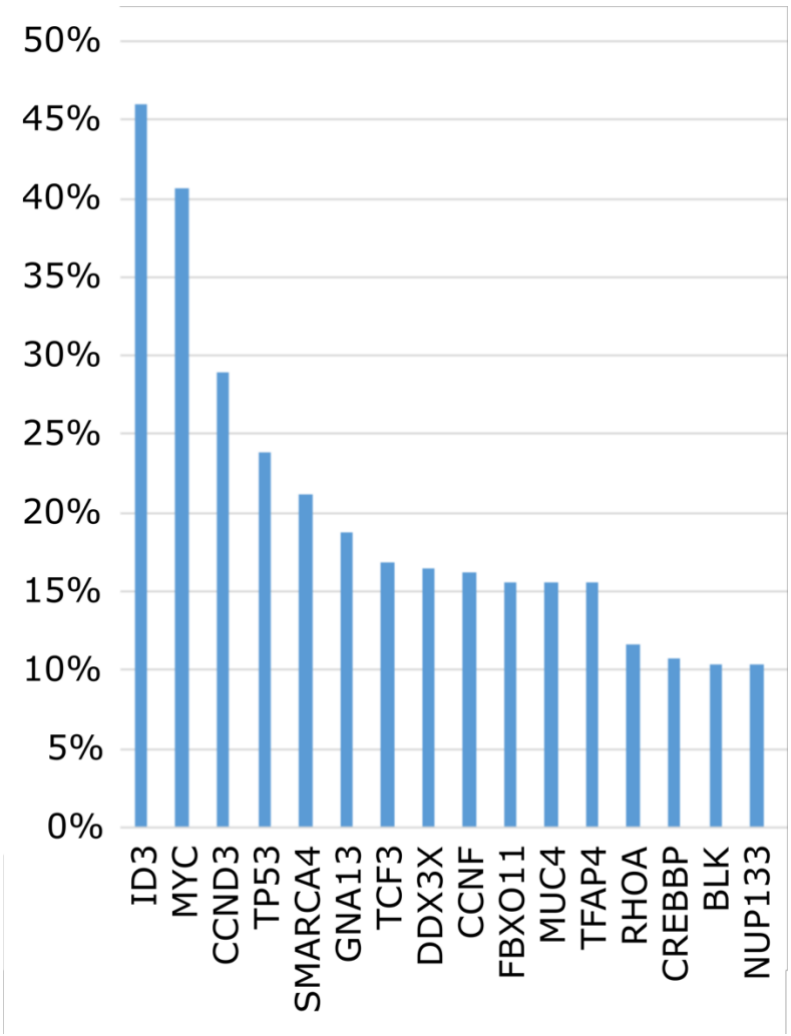


Figure 1.4 - Summary chart displaying the incidence of specific mutations in BL from 6 studies that were identified in over 10% of patients using a combination of Whole Genome Sequencing, Whole Exome Sequencing and Targeted Sequencing approaches.

MYC and *ID3* mutations were the most frequent mutations in BL and were identified in all studies. Mutations in several other genes were identified in >10% of patients, namely *CCND3*, *TP53*, *SMARCA4, GNA13, TCF3* and *RHOA*. Over 70% of BL cases had mutations in either *ID3* or *TCF3*. *ID3* (inhibitor of DNA binding 3) functions by binding E proteins including E2A which is coded by the *TCF3* gene. Silencing mutations of exon 1 of *ID3* were shown to abrogate the ability of ID3 to bind and inhibit E2A and lead to loss of regulation of cell cycle progression in B cells (Richter *et al.*, 2012). Uninhibited activation of *TCF3*/E2A by mutant ID3 or activating mutations of *TCF3* itself leads to activation of PI3K kinase pathway and antigen-independent constitutive activation of the B cell receptor. *TCF3* mutations frequently affect the E47 transcript at the B-HLH (basic helix-loop-helix) domain, removing binding affinity for ID3. *CCND3* encodes cyclin D3 and mutations were frequently

reported in BL that lead to overexpression and protein stabilisation. Additionally, increased *TCF3* expression as a result of *ID3* or *TCF3* mutations can transactivate the *CCDN3* gene leading to deregulation of the G_1 to S phase transition (Figure 1.5) (Schmitz *et al.*, 2012). However, none of these sequencing studies presented associated clinical data to identify prognostic markers.

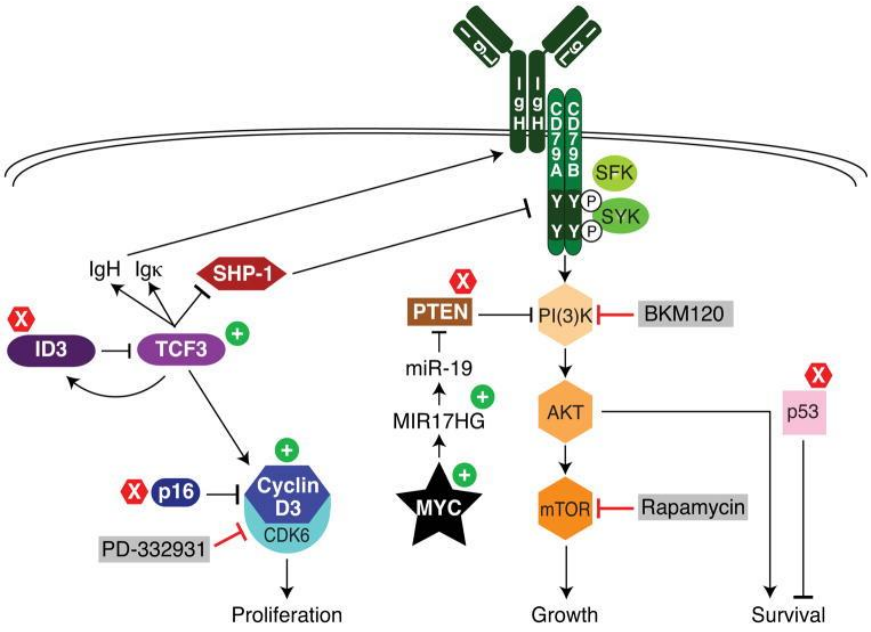


Figure 1.5 - Deregulated pathways in Burkitt Lymphoma. Schmitz et al. *2012.*

Recent studies of sporadic and endemic BL have highlighted genomic events that were enriched in specific subtypes of BL. The primary driver of genetic differences between subtypes is thought to be EBV status (Grande *et al.*, 2019). *AID* activity was significantly higher in EBV positive cases. Regions enriched for non-coding mutations were regularly identified at transcription start sites in EBV positive BL patients that were associated with somatic hypermutation. These genes were the highest expressed in BL tumours of both subtypes, but expression was not associated with mutation at that locus. The same study identified recurrent mutations in apoptotic genes, including *TP53,* in EBV negative BL more frequently than in EBV positive BL, supporting the hypothesis that EBV infection drives inhibition of apoptosis in BL (Grande *et al.*, 2019). Additionally, the incidence of “Burkitt-associated genes” was significantly lower in EBV positive BL than those without EBV involvement. These genetic differences between EBV groups, and strongly associated with tumour subtype, illustrate the need for subtype-specific BL drug targets and prognostic markers.

A study involving whole genome sequencing of 101 BL samples representing all three subtypes reported that the genomic profiles of immunodeficiency-associated and sporadic BL were very similar while endemic BL, particularly the EBV positive cases, were distinct from other subtypes. Mutations in *MYC* were most frequently observed across BL subtypes, with 72 driver genes identified across the cohort (Panea *et al.*, 2019). Genes thought to play roles as oncogenes including *BCL6* and *MCL1* were shown to harbour a series of missense mutations and amplification events while putative tumour suppressor genes such as *DDX3X*, *ARID1A* and *ID3* were repeatedly deleted or harboured truncating mutations. As the first whole genome sequencing study of a large BL cohort, the study reported the novel finding that non-coding events occurred in the majority of driver genes and over 90% of cases had at least one non-coding mutation in a driver gene. For a subset of genes this was proven to be associated with somatic hypermutation, which complements the findings of Grande *et al.* (2019). Further validating their findings, a strong association was observed between EBV positive endemic BL and dysregulation of *AID* function (Panea *et al.*, 2019). Remarkably, *ID3* abrogation was one of the most common events in all three disease subtypes, and investigation in cell lines showed that *ID3* synergised with *MYC* in BL pathogenesis. The authors postulated that due to the similarities in genetic profiles between BL subtypes that clinical trials and new therapeutics identified in sporadic BL will be relevant in the endemic and immunodeficiency-associated subtypes. This may be broadly true, but it is important to remember the significant differences between clinical settings in the Europe/USA compared to sub-Saharan Africa. Vastly different treatment protocols will likely exert unique selection pressures onto the progressing disease, resulting in the acquisition of driver abnormalities which may be associated more strongly with one subtype than the others.

1.3.2 Germinal Centre-like B cell DLBCL (GCB-DLBCL)

When the GCB and ABC subtypes were first reported by Alizadeh *et al.* (Alizadeh *et al.*, 2000), two primary oncogenic events were identified in the GCB subtype; *IGH*/*BCL2* translocation and *REL* amplification. Up to 40% of GCB-DLBCL cases have t(14;18)(q32;q21) *IGH*/*BCL2* translocations which leads to constitutive overexpression of *BCL2*, a feature common to both ABC and GCB-DLBCL subtypes, but more commonly due to *BCL2* amplification or upregulation in ABC-DLBCL.

However high expression of *BCL2* is not solely a result of translocation or amplification, as high expression has been observed in patients with no detectable *BCL2* abnormalities. This would suggest that other mechanisms are necessary for increasing BCL2 expression in these patients, such as NF-κB activation, loss of mir-15 and mir-16 expression (Cimmino *et al.*, 2005) or changes to epigenetic regulation of the *BCL2* gene. Patients with *IGH/BCL2* translocation have been shown to express significantly higher levels of BCL2 and CD10 than their non-translocated counterparts (Iqbal *et al.*, 2004).

Resistance to apoptosis is recognised as one of Hanahan and Weinberg's (Hanahan and Weinberg, 2011) Hallmarks of Cancer. BCL2 plays an important role in the inhibition of apoptosis and as such is known to be important in the pathogenesis of multiple cancer types (Yip and Reed, 2008) after originally being identified by the t(14;18)(q32;q21) translocation in non-Hodgkin Lymphoma. The BCL2 family of proteins consists of pro-apoptotic and anti-apoptotic members that share one or more BH domains (BCL2 homology), and mutation or dysregulation of pro-apoptotic members also occurs in cancer. High BCL2 expression reduces the ability of cytotoxic therapy to kill tumour cells by inhibiting apoptosis. As BCL2 inhibits apoptosis at the mitochondrial membrane by abrogating cytochrome c release, it is able to block apoptotic stimuli originating from almost any source, including the vast majority of cytotoxic drugs utilised in cancer treatment. The t(14;18)(q32;q21) translocation has been shown to be associated with a significantly poorer prognosis in GCB-DLBCL (Figure 1.6) (Visco *et al.*, 2013).

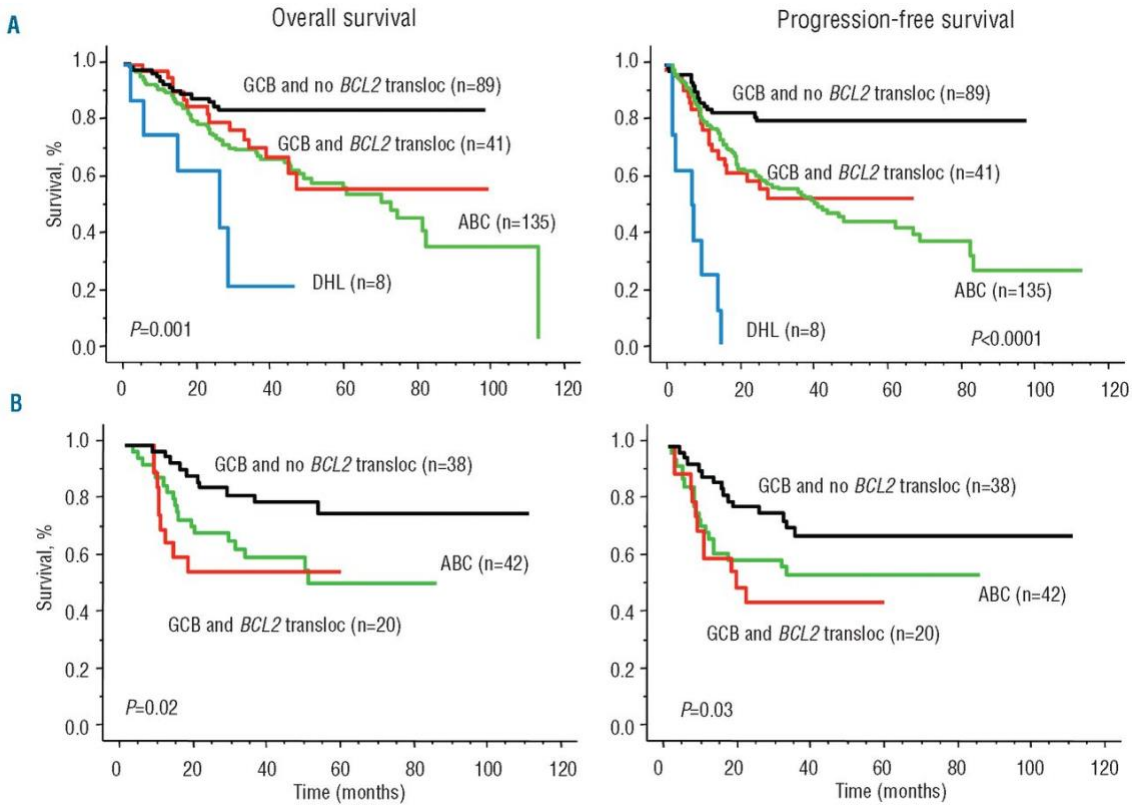


Figure 1.6 - Kaplan Meier plot showing overall and progression-free survival of DLBCL patients with and without BCL2 translocation stratified by GEP defined subgroups.(Visco et al., 2013).

EZH2 (enhancer of zeste homolog 2) is mutated in approximately 20-30% of GCB-DLBCL cases. Wildtype *EZH2* is a histone methyltransferase with a SET domain that plays a role in the Polycomb Repressive Complex 2 (PRC2) and is required for normal embryonic development and VDJ recombination in pre-B cells (Su *et al.*, 2003) before being downregulated. *EZH2*, as part of the PRC2, methylates histone 3 at lysine 27 (H3K27) leading to repression of specific genes. While *EZH2* has important functions in embryonic stem cells, in germinal centre B cells the list of target genes of *EZH2* are not entirely identical, suggesting that the function of the gene in GC cells is different. It has been demonstrated that *EZH2* is required for the formation of the germinal centre and that mutant *EZH2* can cause hyperplasia in the germinal centre (Béguelin *et al.*, 2013). The most common variant of *EZH2* found in GCB-DLBCL is $EZH2^{Y641N}$ which leads to increased methylation of H3K27. It is believed that this change drives growth in B cell lymphoma cells, but the precise mechanism by which this occurs remains elusive. Beguelin *et al.* (2013) also showed that *EZH2*-mutant cell lines have higher levels of proliferation as a result of repression of *CDKN1A*. *CDKN1A* is normally activated in a p53-dependent manner to halt cell cycle progression (Bendjennat *et al.*, 2003). Another role of $EZH2^{Y641N}$ is to

block differentiation of DLBCL cells, just as GCB-DLBCL cells are locked at the GC phenotype. *EZH2*^{Y641N} but not wildtype *EZH2* was shown to do this through the suppression of *PRDM1* (Béguelin *et al.*, 2013), a gene frequently deleted in ABC-DLBCL. Additionally, ChIP-seq experiments showed that other genes important for terminal differentiation including *BCL6*, *HOXA7* and *IRF4*, as well as *PRDM1*, were co-occupied by trimethylated H3K27. Altogether, *EZH2* mutants have the propensity to cause GC B cell hyperplasia, suppress genes required for GC B cells to leave the germinal centre stage of development and provide an ideal state for additional events to transform a GC B cell into a GCB-DLBCL cell.

PTEN (phosphatase and tensin homologue deleted on chromosome 10) gene deletion is observed in about 10% of GCB-DLBCL cases, but loss of PTEN expression has been detected in up to 55% of cases. In an extended cohort of 214 DLBCL cases, 44% of GCB-DLBCL cases were shown to express PTEN by IHC (>5% cells staining) in contrast to 87% in non-GCB-DLBCL (Pfeifer *et al.*, 2013). The results were corroborated by Western blotting in cell lines. Curiously, out of 18 GCB-DLBCL patient samples investigated by quantitative PCR, deletion of *PTEN* was only detected in 3 cases (16.7%), suggesting alternative mechanisms that lead to loss of PTEN expression are involved. Sanger sequencing of the samples revealed an acquired 3bp deletion of *PTEN* in one GCB-DLBCL patient with no PTEN expression. While the incidence of *PTEN* mutation appears to be low, somatic mutations are able to abrogate expression of the gene. *PTEN* functions to negatively regulate the PI3K/AKT pathway. The PI3K/AKT pathway is implicated in DLBCL as a driver of proliferation and survival of tumour cells and is activated by phosphatidyl-inositol-3,-4,-5-triphosphate. PTEN dephosphorylates this molecule, thus abrogating the pathway. Loss of functional *PTEN* leads to the pathway being activated and supports tumour growth. This has been supported by detection of an inverse correlation between PTEN expression loss and levels of phosphorylated AKT (Pfeifer *et al.*, 2013) and the discovery that PTEN-deficient GCB-DLBCL cells respond well to PI3K inhibitor treatment. Outside of the PI3K/AKT pathway, PTEN is important for genome stability through upregulation of *RAD51* to prevent double strand breaks (Shen *et al.*, 2007).

MYC translocation is the hallmark of BL but also occurs in GCB-DLBCL patients at lower incidence. These patients have a significantly inferior prognosis compared to their *MYC*$^{-}$ counterparts and have been associated with a higher proliferation rate

with 50% of *MYC*$^{+}$ having a Ki67 above 90% compared to 7.5% of *MYC*$^{-}$ cases (Savage *et al.*, 2009).

Amplification of the *REL* gene, coding for the REL proto-oncogene of the NF-κB transcription factor family, has been reported in approximately 30% of GCB-DLBCL cases (Scholtysik *et al.*, 2015). *REL* plays an important role in normal lymphocyte biology as it is the predominant NF-κB dimer after B cell maturation (Table 1.2) (Gilmore and Gerondakis, 2011). B cell division and survival in response to stimuli is vital to initiation of an adaptive immune response and REL is required by B cell receptor activated B cells to progress from G1 to S phase of the cell cycle (Grumont *et al.*, 1998), as well as for *MYC* upregulation in G1. c-Rel also plays an anti-apoptotic role in BCR activated cells by activating Bcl-2 family proteins A1 and Bcl-XL, each playing a different role in B cell survival as evidenced by their different response kinetics to c-Rel activation. c-Rel is also required for successful isotype switching from IgG to IgE (Kontgen *et al.*, 1995). Despite NF-κB expression being highest in ABC-DLBCL, *REL* amplification is more frequently reported at higher incidences in GCB-DLBCL (Scholtysik *et al.*, 2015) and as a result it has been suggested that *REL* amplification is required early in the development of GCB-DLBCL, but not in disease progression. Despite the high incidence of *REL* amplification in DLBCL, no DLBCL cell lines exist with the abnormality. It is possible therefore that *REL* amplification aids growth of GCB-DLBCL cells *in vivo* only, and additional unknown factors are required for *in vitro* cell growth.

GENE/PROTEIN	PROTEIN FUNCTION
CELL PROLIFERATION/CELL GROWTH	
C-REL	Transcription factor
C-MYC	Transcription factor
IRF-4	Transcription factor
E2F3A	Transcription factor
EP300	Histone acetyltransferase
CD21	Complement receptor
CD40	Cell surface receptor
SFN	14-3-3 protein
GM-CSF	Hematopoietic growth factor
TGFß	Growth factor
IL-2	Cytokine
IL-4	T cell cytokine
APOPTOSIS/CELL SURVIVAL	
BCL-2	Anti-apoptotic
BFL-1/A1	Anti-apoptotic
BCL-XL	Anti-apoptotic
MIR-21	Pro-apoptotic for ß cells
ADHESION/CELL ARCHITECTURE	
ICAM-1	Cell adhesion
SELECTIN	Cell adhesion; binds sugars
MMP-1	Metalloproteinase
EPHB2	Receptor tyrosine kinase (repressed)
IMMUNE CELL FUNCTION	
GAMMA1	Ig heavy chain
GAMMA4	Ig heavy chain
TNF-A	Cytokine
IL-12	p35 cytokine
IL-13	Cytokine
IL-21	Cytokine
IL-23	Cytokine
CD40L	CD40 ligand
BLYS/BAFF	TNF-like cytokine
IP-10	Chemokine
LIGP1	GTPase
MIG	Macrophage cytokine
FOXP3	Transcription factor
DNA REPAIR/DAMAGE	
ATM	Protein kinase
CLASPIN	Cell cycle kinase
SKP2	S-phase kinase- associated factor

Table 1.2 - Partial List of c-Rel targets - Adapted from Gilmore and Gerondakis, 2011.

Genomic aberrations in a large DLBCL cohort of 148 cases were investigated by Scholtysik *et al*. (2015) using high-resolution Affymetrix Mapping 250k Sty2 SNP arrays. Figure 1.6 displays the genomic aberrations identified by the investigation for all DLBCL cases pooled, ABC subtype only, GCB subtype only and a comparison between subtypes. As might be expected due to the similar cell of origin of BL and GCB-DLBCL, abnormalities identified in BL such as gain of 1q, gain of 7 and abnormalities of 13q were common in GCB-DLBCL (Scholtysik *et al.*, 2015). Additionally, GCB-DLBCL harboured copy number changes such as 2p15 amplification involving the *REL* gene and 6q loss involving *PRDM1* (Figure 1.7).

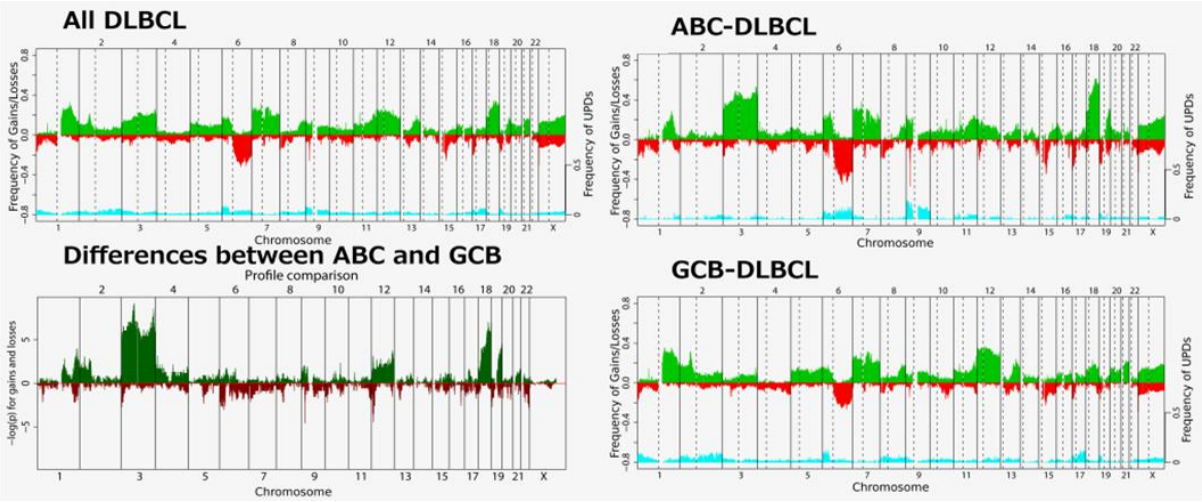


Figure 1.7 - Genomic Abnormalities detected by high-resolution SNP arrays in DLBCL, separated by subtype. Adapted from Scholtysik et al. *2015.*

1.3.3 Activated B cell-like Diffuse Large B cell Lymphoma (ABC-DLBCL)

The cell of origin of ABC-DLBCL is thought to be plasmablastic B cells that have not left the germinal centre, based on gene expression profiling (Alizadeh *et al.*, 2000). The block on differentiation of these cells gives them an activated B cell phenotype without fully differentiating to plasma cells. This may be due to several different mechanisms. *BCL6* (B Cell Lymphoma protein 6) is highly expressed in GC B cells but not in activated plasma B cells. Differentiation to plasma cells requires BLIMP1, coded by the *PRDM1* gene on chromosome 6q and BCL6 and BLIMP1 mutually inhibit each other, so for successful differentiation of a GC B cell to a plasma B cell downregulation of BCL6 is required. BCL6 expression is dependent on a number of factors, most notably *BCL6* mutation, CD40 signalling, growth factors including IL-21 and STAT5, and signalling from the B cell receptor. *BCL6* translocations with various

partners are present in approximately one quarter of ABC-DLBCL cases leading to deregulation of the gene, but has not been shown to correlate with expression (Iqbal *et al.*, 2007).

ABC-DLBCL cases are much more dependent on NF-κB signalling than GCB-DLBCL. Constitutive activation of NF-κB is a hallmark of ABC-DLBCL, leading to cell proliferation and resistance to apoptosis, and therapeutic targeting of IKK has been shown to kill ABC-DLBCL but not GCB-DLBCL cells (Lam *et al.*, 2005). shRNA screening showed that NF-κB activation in ABC-DLBCL is dependent on the CARD11, BCL10 and MALT1 (CBM) complex (Ngo *et al.*, 2006). This complex is important for IKK activation in normal biology, but in ABC-DLBCL the pathway is activated by alternative methods including *CARD11* mutation. Constitutive activation of IKK leads to abnormally high levels of IκB ubiquitination and degradation and as a result nuclear translocation of NF-κB family members. CARD11 (caspase recruitment domain-containing protein 11) is a scaffold protein for NF-κB and is expressed in B cells. Missense mutations have been detected in approximately 10% of ABC-DLBCL cases (Lenz *et al.*, 2008). Additional CBM-independent mechanisms of NF-κB activation have been identified including deletion or inactivation of A20 (Kato *et al.*, 2009). A20 is a protein that acts to negatively regulate NF-κB family members and is located at 6q23.3. A20 induction occurs as a result of stimulation by TNF-α (Dixit *et al.*, 1990).

Figure 1.6 displays the results of copy number analysis of 148 DLBCL cases, including 49 ABC-DLBCL. GISTIC, a statistical tool used for detecting significantly aberrant genomic regions in cancer, revealed 13 recurrent gains and 25 recurrent deletions in ABC-DLBCL (Scholtysik *et al.*, 2015). Deletions of chromosome 6 were most common in the ABC subtype and three minimal regions were identified by GISTIC: 6p21.23, 6q21 and 6q23.3. 6p21.23 harbours the HLA region which is commonly variable by copy number analysis. *PRDM1* resides at 6q21, and deletion of this region would support existing evidence that PRDM1 is required for terminal differentiation to plasma cells and that ABC-DLBCL cells are blocked before this stage. Deletion at 9p21.3 involved the two genes *CDKN2A* and *CDKN2B*, both of which are involved in cancer-related pathways. *CDKN2A* codes for two transcripts, $p16^{INK4}$, which is a MDM2 binding protein, and $p14^{ARF1}$ which is a CDK inhibitor.

CDKN2B encodes $p15^{INK4B}$, which has been shown to inhibit cell growth in response to TGF-β (Stone *et al.*, 1995).

Additional differences between ABC- and GCB-DLBCL subtypes detected by array were a higher incidence of gains of chromosomes 3 and 18 in the ABC group and gain of chromosome 18 in GCB-DLBCL. Chromosomes 3 and 18 harbour the *BCL6* and *BCL2* genes respectively and the difference in BCL2 copy number gain frequency reflects previously reported data as translocations are the primary cause of overexpression of BCL2 in GCB-DLBCL and would not be detectable by SNP array.

1.3.4 Paediatric DLBCL

Paediatric DLBCL makes up approximately 10% of non-Hodgkin lymphomas diagnosed in patients aged up to 14 years in the UK (HMRN, 2019). Paediatric DLBCL is almost exclusively of the GCB subtype (Iqbal *et al.*, 2004). However unlike adult GCB-DLBCL the t(14;18) translocation that places the *BCL2* gene under control the IGH promoter is extremely rare in paediatric patients (Oschlies *et al.*, 2006) and BCL2 expression is less common. In contrast, *IRF4*/*IGH* translocations have been identified as more common in paediatric DLBCL cases (Salaverria *et al.*, 2011). Few studies exist examining the molecular characteristics of paediatric DLBCL. The Poirel study of cytogenetic data collected within the FAB/LMB96 trial highlighted *MYC* rearrangement was more frequent than reported in adults at approximately 33% of paediatric cases compared with 10% in adults (Poirel *et al.*, 2009). A recent study attempted to characterise the paediatric disease using gene expression and CGH microarray analysis, FISH and IHC (Klapper *et al.*, 2012). They reported that amplifications of 2p16 (*REL*), a common abnormality in adult DLBCL, were not detected in the childhood cases. Elucidation of a clear cut-off between paediatric and adult lymphoma was then attempted using known molecular markers. By plotting the incidence of these abnormalities against patient age (Figure 1.8), visual inspection suggested that many of these abnormalities differ in frequency with age. Rather than

a stark cut-off between age groups, a spectrum of continuous change was suggested, with the most notable transition observed between ages 24 and 36.

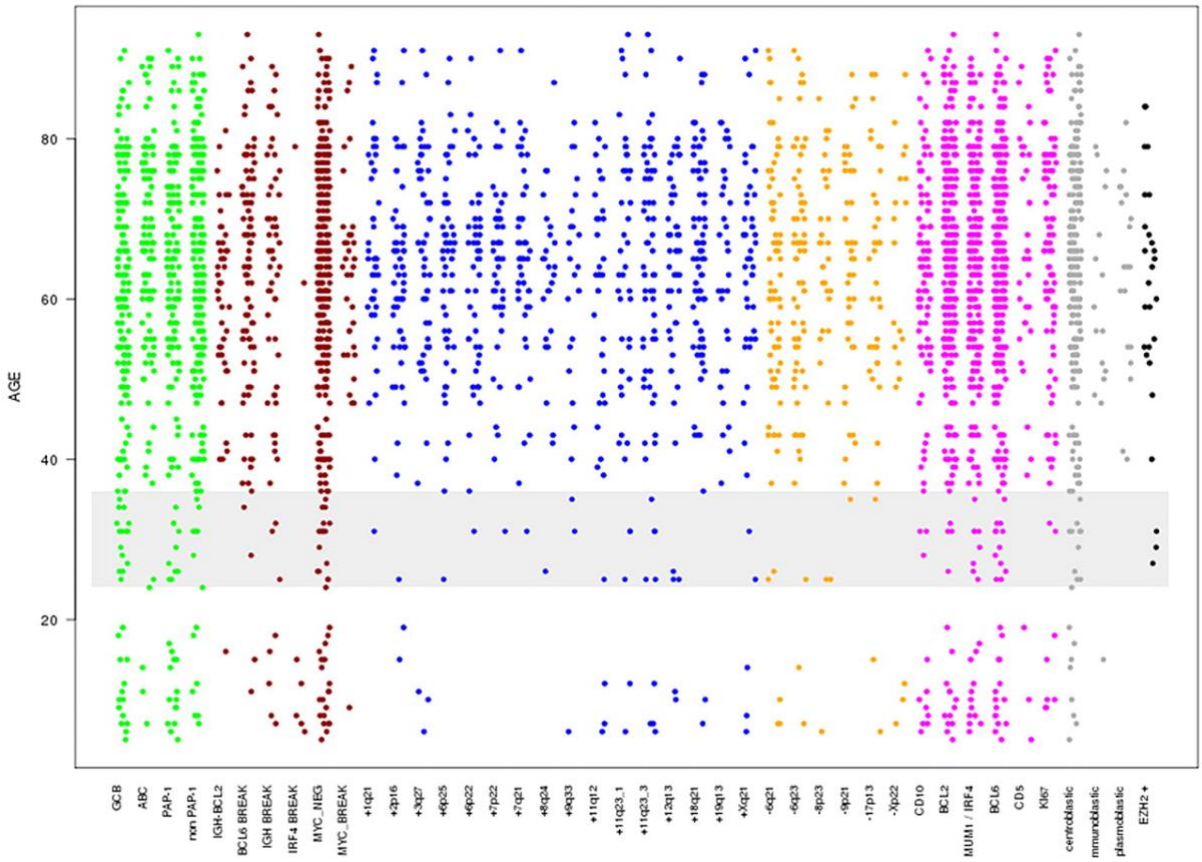


Figure 1.8 - Scatter plot displaying incidence of abnormalities against patient age for aggressive B cell lymphomas other than Burkitt Lymphoma. Klapper et al. *2012.*

Statistical analysis confirmed that the significantly superior prognosis of paediatric DLBCL compared to the adult form was independent of GCB/ABC subtype bias and that age was an independent factor for predicting outcome. BCL2 expression by IHC and an ABC-type gene expression signature were separately confirmed to be prognostic factors independent of age. The results of the study strongly suggest there is no biological rationale for the current dichotomous treatment strategies. It was demonstrated that the paediatric disease differs in its molecular profile to the adult disease, but that it is not currently possible to determine an age cut-off between the two to guide therapy, and an “age evolution model” was suggested instead. This leaves a therapeutic no-mans-land of teenagers and young adults who are insufficiently treated by adult regimens and may benefit from more aggressive chemotherapy regimens administered to paediatric cases.

1.4 Current treatment

Treatment of BL and DLBCL in patients below the age of 18 is routinely according to the FAB/LMB96 trial protocol which delivers intensive multi-agent chemotherapy including cyclophosphamide, vincristine, prednisolone, doxorubicin, methotrexate, cytarabine and etoposide (Patte *et al.*, 2007). Within this regimen patients are further stratified into group A, B or C. Group A comprises completely resected stage I and abdominal stage II patients, who are considered to be at the lowest risk and receive just two courses of COPAD chemotherapy. Group C patients have stage IV disease or Burkitt's ALL (L3ALL), as well as patients with CNS involvement and represents the most advanced stage patients. Group B encompasses all patients not eligible for groups A or C. Group B and C patients receive an induction phase of two courses of COPADM, followed by a consolidation phase of two courses. Group B patients receive two CYM consolidation courses and stop. Group C patients receive two courses of high dose cytarabine and etoposide, followed by two maintenance courses. Disease status is reassessed following CYM1 (group B patients) or CYVE2 (group C patients) and any residual mass >2cm must be biopsied/resected to confirm histological remission. Failure to confirm histological remission results in escalation of therapy. For adults with Burkitt lymphoma, those who are adequately fit will receive a protocol similar to FAB/LMB96 – R-CODOX-M-IVAC. For frailer adults not able to tolerate the intensity of R-CODOX-M-IVAC, an alternative strategy will be used.

The recent Inter-B-NHL-Rituximab trial established in 2010 has significantly improved the outcome of paediatric B-NHL patients. Patients in group B with more than double the upper normal range of LDH are now given rituximab upfront at the point of initial diagnosis, as well as all group C patients. This has led to an event-free survival of over 94% in paediatric BL and DLBCL cases (Minard-Colin *et al.*, 2016).

Treatment for DLBCL in the UK (as well as in the Klapper *et al*. study of German patients (Klapper *et al.*, 2012) is uniform across subtypes for adults. R-CHOP is a regimen of chemotherapy combined with the anti-CD20 antibody Rituximab. CHOP has been shown to be the gold standard in first-line treatment of DLBCL in adults (Messori *et al.*, 2001). However, it is associated with a degree of toxicity to the patient and is difficult for some older patients to tolerate. One study of 135 aggressive NHL patients receiving standard CHOP reported that 27 patients (20%)

suffered congestive heart failure within 1 year of treatment, three of whom died as a result (Limat *et al.*, 2003). Additionally, approximately 40% of patients treated with R-CHOP will go on to relapse or their disease will become refractory (Sehn, 2012). Therefore there is a clear clinical demand for targeted therapies that manipulate the biology of B cell lymphoma to improve patient outcome and reduce the debilitating toxicity that current approaches cause (Minard-Colin *et al.*, 2016).

1.5 Development of chemotherapy resistant B-NHL

Cancer is known to progress and evolve in a similar manner to Darwinian evolution. It has recently been postulated that disease progression and treatment resistance arise as a result of clonal selection from an inherently heterogenous tumour (intrinsic resistance), or ongoing genomic evolution of the tumour reacting to selection pressures caused by frontline therapy, or indeed a combination of the two methods (Asic, 2016). Intrinsic resistance arises early during tumorigenesis, while acquired resistance occurs on or after treatment. Studying the genomic changes throughout the development of disease progression may provide clinical solutions to intrinsic resistance. Acquired resistance is a more complicated clinical challenge, as it has been shown that targeted therapies can drive malignancies to acquire resistance to the targeted agent (Asic, 2016). In fact, most FDA-approved targeted therapies used in cancers not driven by a single, dominant oncogene lead to resistance to therapy within one year of salvage therapy (Ellis and Hicklin, 2009). This is particularly prevalent in DLBCL, where it has been shown that rituximab treatment reduces salvage rates of patients who go on to receive autologous stem cell transplants if they relapsed within 12 months of frontline therapy (Gisselbrecht *et al.*, 2010).

In BL, relapse occurs much quicker than in DLBCL patients, particularly in paediatrics, with relapse usually occurring within one year of diagnosis. Rapid recurrence allows less time for the tumour population to diversify, suggesting that the pattern of evolution in BL disease progression is likely to be linear, with late divergence of clones (Juskevicius *et al.*, 2017). A second model of clonal evolution in B-NHL was recurrently seen in DLBCL where there are genomic aberrations in common between the diagnostic clone and the major clone at disease progression, but that the clones diverged early, representing a more “branched” pattern of evolution. Both patterns are seen in DLBCL, with late-divergent/linear evolution occurring most frequently (65%) and early divergent/branching progression occurring

in approximately 30% of cases (Juskevicius *et al.*, 2016). Both patterns of clonal evolution are summarised in Figure 1.9.

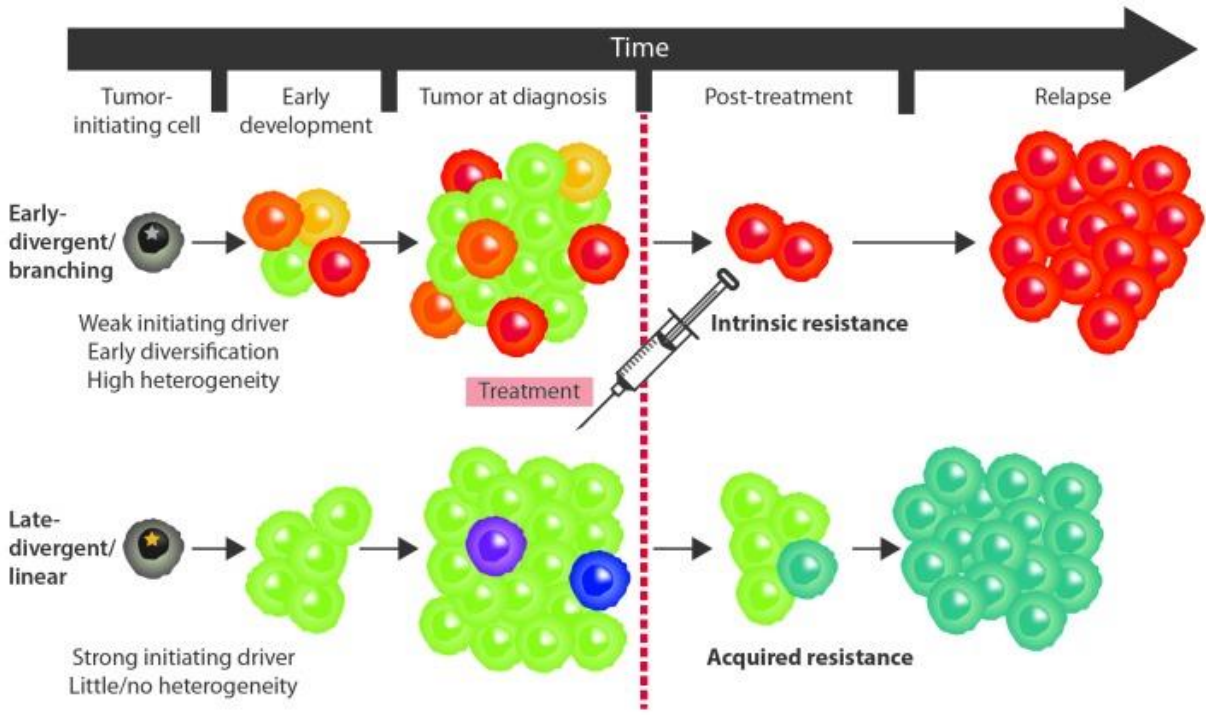


Figure 1.9 - A cartoon demonstrating the two patterns of evolution observed in DLBCL. Adapted from Juskevicius et al. (2017)

The only study of sequential karyotypes from BL relapse/refractory patients was performed using traditional karyotyping methods (Aukema *et al.*, 2015). The investigation revealed that while BL has a relatively quiet genomic landscape at diagnosis, there was a significant increase in the number of cytogenetic aberrations identified in sequential samples (Aukema *et al.*, 2015) (Figure 1.10). Karyotyping of tumour cells is a very low-resolution technique that may have significantly under-represented the complexity of tumours at both diagnosis and relapse.

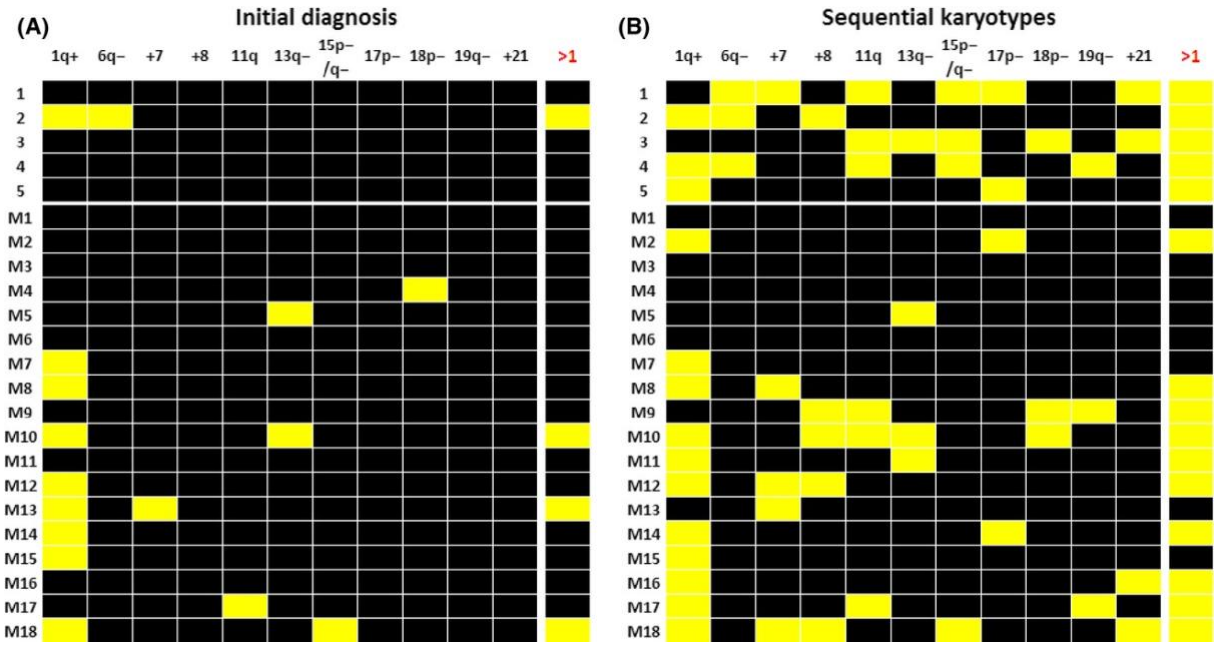


Figure 1.10 - A heatmap of recurrent secondary abnormalities in (A) diagnostic and (B) sequential BL samples.

In this study, the pattern of evolution was shown to be primarily linear in 20/25 BL, but more studies are needed to better understand the pattern of clonal evolution in this malignancy. It was also reported that clonal aberrations seen in sequential samples were present in a subclonal population of cells studied at diagnosis, particularly trisomy 12 and dup(1q). Mathematical modelling of BL relapse supported the presence of primarily early linear relapses, and showed that even later relapses were primarily linear (Aukema *et al.*, 2015). To better unravel the clonal evolution of paediatric B-NHL it is important to identify putative prognostic genomic markers. Evaluation of these prognostic events in paediatric B-NHL may reveal genomic drivers of intrinsic therapy resistance. Alternatively, these markers may simply drive genomic instability in the tumour and lead to an increased likelihood of acquired resistance arising.

1.6 Aims and Hypothesis:

Despite improvements in outcome of patients diagnosed with aggressive B-NHL, relapsed and refractory disease remains a major clinical challenge. The failure to fully respond to treatment suggests a fundamental difference in the underlying biology of relapsed/refractory disease. The central hypothesis for this project is that by interrogating the largest UK dataset of B-NHL patients it will be possible to identify genomic abnormalities which drive resistance to chemotherapy and highlight genes and pathways that could be used to stratify patients into clinically significant subgroups or targeted by therapeutics to improve patient outcome and overcome resistance to chemotherapy.

Current research into aggressive B cell lymphoma and identification of prognostic markers is regularly hindered by small sample sizes and insufficient clinical information. The majority of published studies focus on adult B-NHL and little is known about BL and DLBCL in children and adolescents. This analysis will reveal which abnormalities promote survival of malignant cells through current intensive chemotherapy regimens and remain in the resistant tumour at point of disease progression. The patient cohort will form the basis of this PhD project and will be integrated with published datasets in order to:

1. Model paediatric B-NHL by integrating copy number arrays from in-house datasets and data from the literature to identify putative prognostic markers.
2. Further identify and characterise potential prognostic markers with integrated whole-exome sequencing analysis, evaluating the clinical relevance of identified markers.
3. Investigate diagnostic and relapse paired samples to investigate the clonal evolution of disease progression.

The hypothesis of this thesis is that high resolution copy number array analysis when compared with exome sequencing data will identify prognostic markers in paediatric B-NHL. These markers can then be further validated in a trial cohort and used to improve outcomes in paediatric B-NHL patients.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Laboratory Equipment

Equipment	Manufacturer
Precellys Homogeniser	Bertin
Thermobrite Hybridiser	Fisher Scientific
SureCycler 8800 Thermal Cycler	Agilent
Water Bath	Grant Instruments
Coplin Jars	Thermofisher Scientific
Qubit 3.0 Fluorometer	Thermofisher Scientific
Nanodrop 2000	Thermofisher Scientific
DNA Gel Tank	Engineering and Design Plastics Ltd.
Electrophoresis Powerpack	BioRad
ChemiDoc Visualiser	BioRad
Tapestation 4200	Agilent
Bioanalyzer 2100	Agilent
Fluorescence Microscope (BX61)	Olympus

2.1.2 Software

Copy Number Analysis software	Developer
Nexus Copy Number 10.0	Biodiscovery
Integrative Genome Viewer (IGV)	Broad Institute
Oncoscan Console	Affymetrix
GISTIC	Broad Institute
GISTIC2.0	Broad Institute

Sanger Sequencing Analysis	Developer
FinchTV	Geospiza Inc.
BLAT	Ensembl, EMBL-EBI

Whole exome sequencing	Developer
GATK3.8	Broad Institute
Samtools	Heng Li (Li *et al.*, 2009)
FastQC	Babraham Institute
Bedtools	Quinlan Lab (Quinlan and Hall, 2010)
Variant Effect Predictor	Ensembl, EMBL-EBI

Statistical Analysis	Developer
RStudio	RStudio Inc.
Excel 16	Microsoft

R packages	Developer/URL
survival	https://github.com/therneau/survival
devtools	https://github.com/r-lib/devtools

survminer	https://github.com/kassambara/survminer
tidyr	https://github.com/tidyverse/tidyr
dplyr	https://github.com/tidyverse/dplyr
coxrt	https://github.com/Bella2001/coxrt
ggplot2	https://github.com/tidyverse/ggplot2
lmtest	https://cran.r-project.org/web/packages/lmtest/index.html
maftools	https://github.com/PoisonAlien/maftools
genvisr	https://github.com/griffithlab/GenVisR
forestplot	https://github.com/gforge/forestplot

GitHub Repository	**URL**
Scripts written for use in this thesis	https://github.com/anewman66/phd-thesis-2020
Matt Bashton's GATK pipeline	https://github.com/MattBashton/MB-GATK-SGE
GISTIC2.0 Dockerfile	https://github.com/sandertan/docker-gistic2

2.1.3 Chemicals and reagents

Reagent	**Supplier**
Fixogum Rubber Cement	Generon
VectaShield Mounting Medium with DAPI	Vectorlabs
Thermo Scientific dNTP kit	Fisher Scientific
GelRed	VWR
Hyperladder (100bp)	Bioline
Electrophoresis-grade agarose	Fisher Scientific
20X SSC	Fisher Scientific

2.1.4 Experimental kits

Kit	**Supplier**
Qiagen AllPrep DNA/RNA/Protein extraction kit	Qiagen
Qiagen AllPrep DNA/RNA/miRNA extraction kit	Qiagen
QIAquick PCR Purification kit	Qiagen
Cytocell Tissue Pre-treatment kit	Cytocell
Thermo Scientific Qubit Broad Range DNA Assay	Fisher Scientific

Agilent RNA 6000 Nano kit	Agilent
Agilent RNA ScreenTape for Tapestation	Agilent
Applied Biosystems AmpliTaq Gold™ DNA Polymerase with Buffer II and MgCl2	Life Technologies
Thermo Scientific dNTP set	Fisher Scientific
Ambion DNA-free Turbo DNAse kit	Life Technologies
Precellys Ceramic kit 2.8mm 50x2 ml tubes, prefilled with ceramic beads	Peqlab

2.1.5 Oligonucleotide sequences

Primer oligonucleotide sequences for *TP53* exons 5-8 are presented in Table 2.1.

Exon	**Forward sequence**	**Reverse sequence**
Exon 5	CTCTGTCTCCTTCCTCTTCC	GCAATCAGTGAGGAATCAGAGG
Exon 6	GCCTCTGATTCCTCACTGAT	GGAGGGCCACTGACAACCA
Exon 7	AGGCGCACTGGCCTCATCTT	CAGGGGTCAGAGGCAAGCAGA
Exon 8	GAGCCTGGTTTTTTAAATGG	TTTGGCTGGGGAGAGGAGCT

Table 2.1 - Primer sequences for TP53 exons 5-8 used for Sanger sequencing.

2.1.6 FISH Probes

Probe	Supplier
IGH/c-MYC Plus Translocation, Dual Fusion Probe	Cytocell
IGL/cMYC Dual Fusion Probe	Cytocell
IGK/cMYC Dual Fusion Probe	Cytocell
Aquarius Pathology FISH MYC breakapart probe	Cytocell

2.2 Approvals

2.2.1 Ethics approval for studies using CCLG Biobank tumour material

Tumour samples were collected from patients diagnosed with paediatric lymphoma with informed written consent obtained from each participant or their parent/legal guardian as per CCLG (Children's Cancer and Leukaemia Group) guidelines and protocols at local centres. Original consent forms are held by the CCLG. Fresh frozen tumour samples were stored in accordance with the Human Tissue Act (2006). Project and ethics approval were granted via the CCLG Biological Studies Steering Group (REC 18/EM/0134; Biological Study 2012 BS 08).

2.2.2 Ethics approval for studies using patient tumour FNAs from Queen Elizabeth Central Hospital, Malawi

Tumour samples were collected from patients diagnosed with suspected lymphoma at QECH, Blantyre, Malawi, between October 2009 and March 2014. This study was approved by the Malawi College of Medicine Research and Ethics Committee (COMREC:P.07/12/1249) and the Newcastle University Ethics Committee (3314/2016) and written informed consent for the use of tumour samples for research was obtained.

2.3. Methods:

2.3.1 Patient Cohorts

Distinct patient cohorts were used for each chapter, with some overlap of patients analysed between chapters. There were three primary sources of patient samples and data: CCLG cohort, literature search of data repositories and Malawi cohort. Summaries detailing the patients of each cohort involved in chapter of results are detailed in Figures 2.1-2.5.

Analysis in Chapter 3 was performed on patients from the CCLG biobank with existing copy number array data generated on the Affymetrix Genome-Wide Human SNP6.0 array platform and cases identified from the literature (Figure 2.1). Raw copy number array data from seven studies identified via the Gene Expression Omnibus (GEO) data repository was downloaded for analysis. Additional segmented data from one study was taken from the supplemental data of a paper in the British Journal of Haematology (Havelange *et al.*, 2016b).

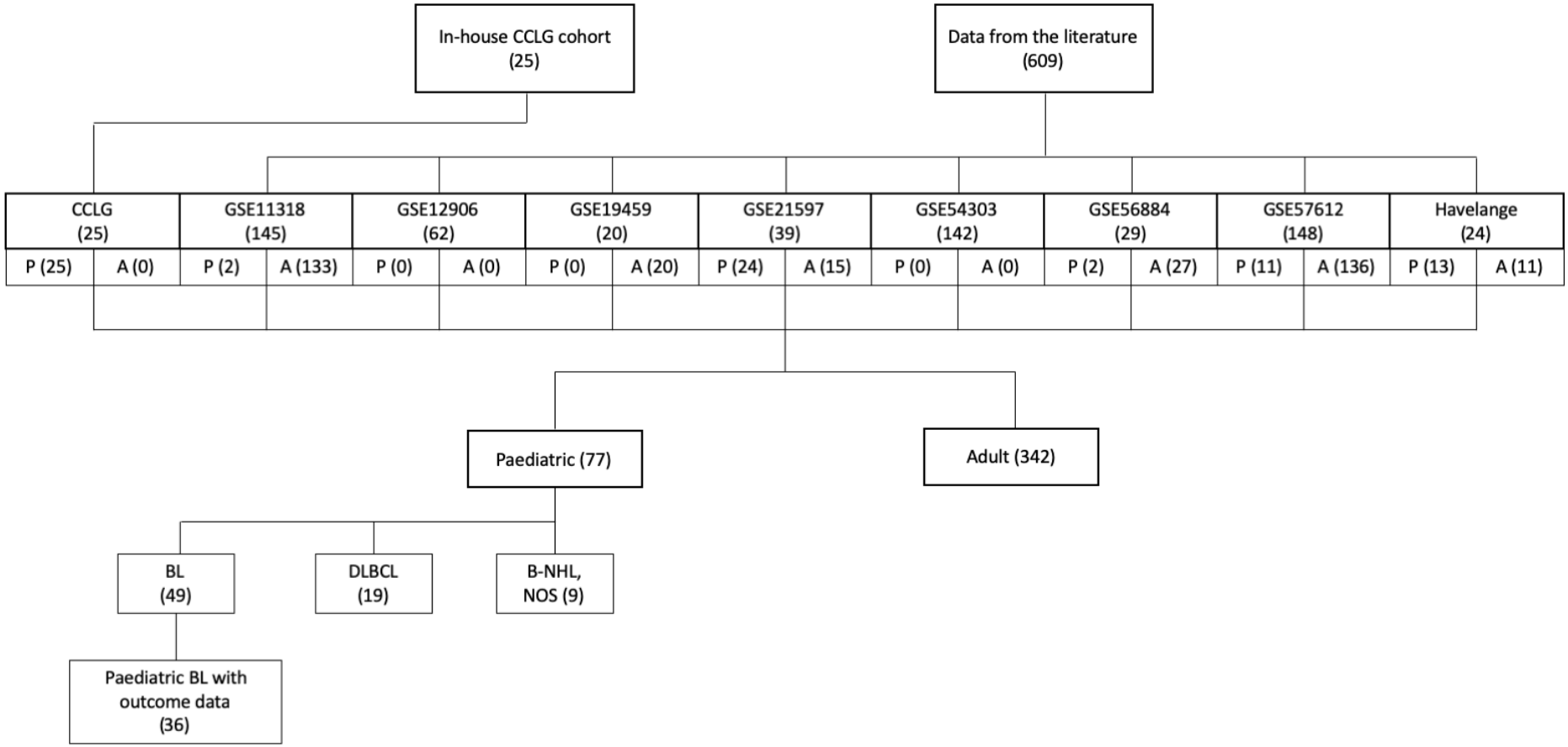


Figure 2.1 - Consort diagram detailing the makeup of the cohort used for the Pilot Project described in Chapter 3. Some data from some accession numbers had one or more cases with no associated age data: GSE11318: 10 cases (6.9%), GSE12906: 62 cases (100%), GSE54303: 142 cases (100%), GSE57612: 1 case (0.7%).

The patient cohort of 162 cases used in Chapter 4 is detailed in Figure 2.2. Running parallel to the copy number analysis in Chapter 3, new samples were prepared for copy number analysis through DNA extraction and QC. DNA samples were sent for copy number array hybridisation to either Affymetrix Cytoscan HD arrays or Affymetrix Oncoscan v2 MIP arrays. If sufficient DNA was available for the 25 CCLG cases with existing Affymetrix Genome-Wide Human SNP.60 array data then the samples were re-analysed on the Affymetrix Cytoscan HD array kit.

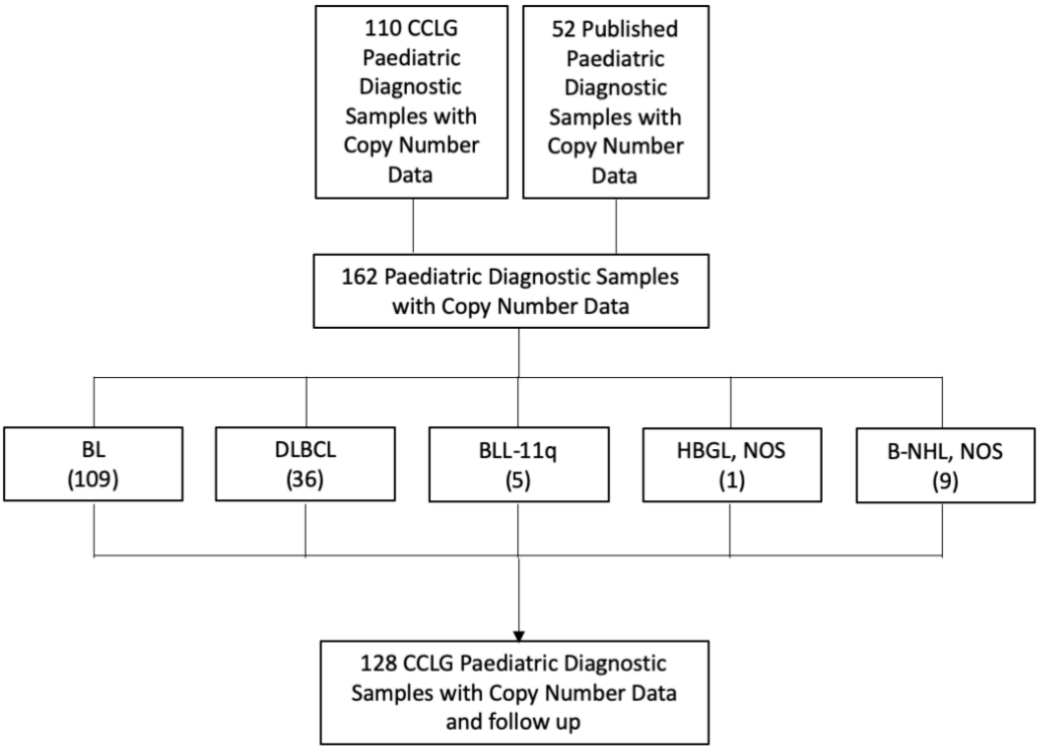


Figure 2.2 - Consort diagram detailing the makeup of the cohort used for analysis in Chapter 4.

The patient cohort analysed in Chapter 5 was made up of a subset of the cohort from Chapter 4. 95 patients with both copy number array data and *TP53* sequencing data were selected for analysis. Cases from the literature were excluded from analysis as large numbers of these cases were lacking mutation data. There were 15 cases without exome sequencing data that could not be analysed by Sanger sequencing due to a lack of DNA and these cases were excluded from the analysis. The breakdown of patients involved in the analysis is detailed in Figure 2.3.

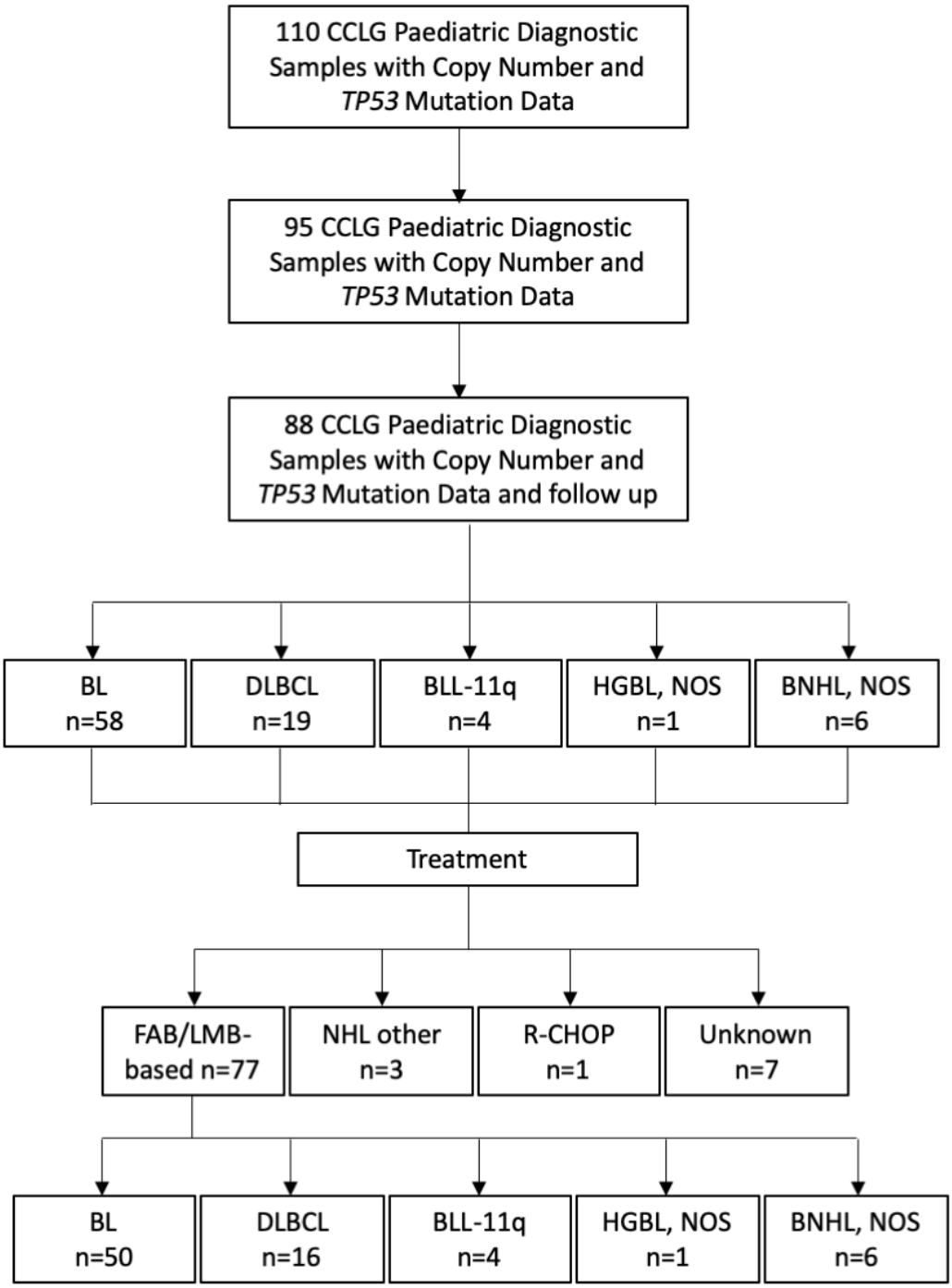


Figure 2.3 - Consort diagram detailing the makeup of the CCLG cohorts analysed in Chapter 5.

Two patient cohorts were analysed in Chapter 6 to compare the endemic and sporadic BL. The sporadic cohort was comprised of a subset of BL patients from the cohort analysed in Chapters 4 and 5 with four additional locally diagnosed cases added. The four additional cases were diagnosed and treated at the Great North

Children's Hospital at the RVI, Newcastle. Tumour DNA from the patient samples was kindly supplied by Dr Simon Bomken.
Endemic BL patients were acquired from a collaboration with the Queen Elizabeth Central Hospital in Blantyre, Malawi. The makeup of both cohorts are detailed in Figure 2.4.

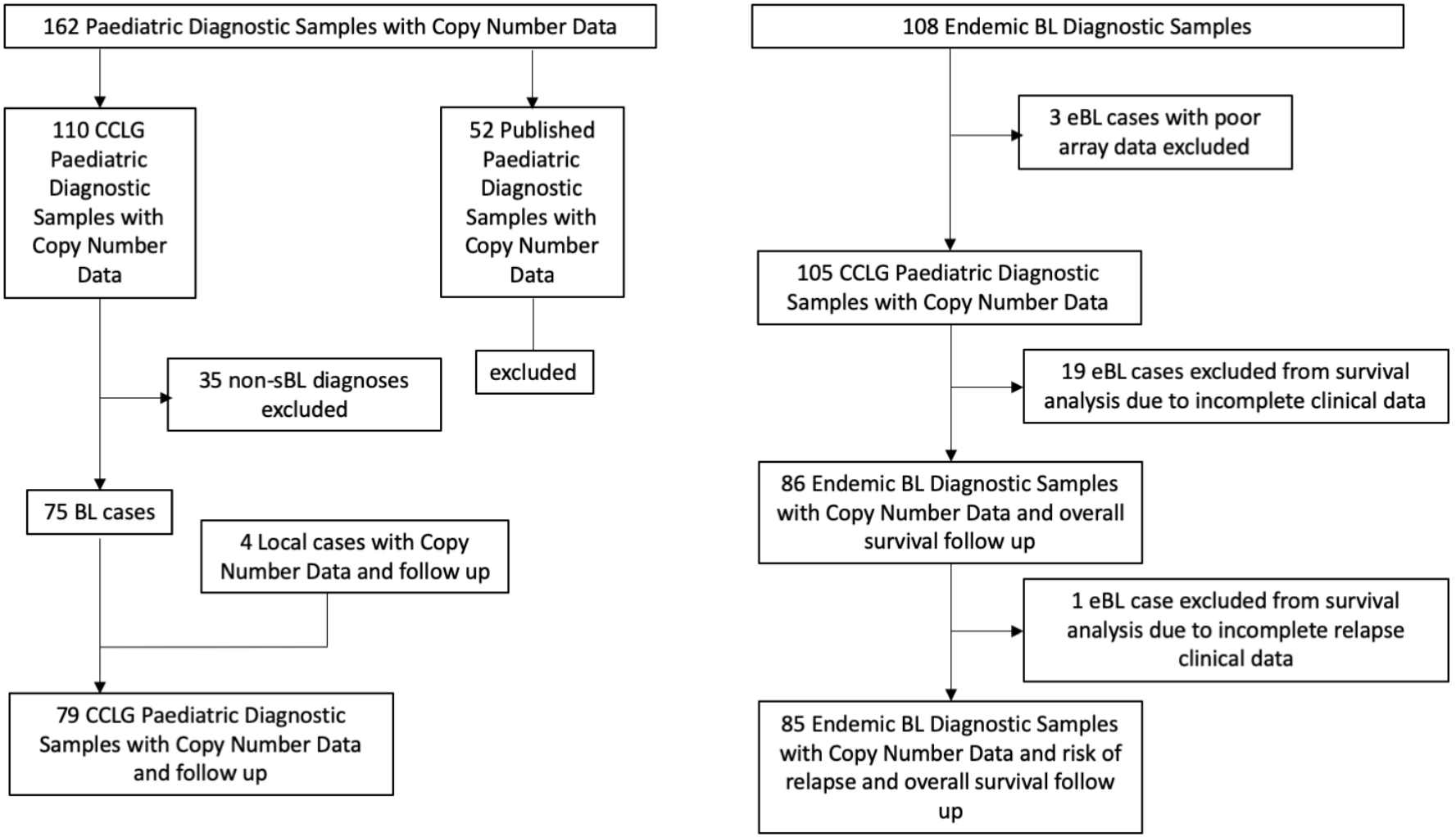


Figure 2.4 - Consort diagram detailing the makeup of the sporadic BL and endemic BL cohorts used in for analysis in Chapter 6.

11 patients who underwent disease progression were analysed in Chapter 7. Nine patients were from the CCLG cohort and copy number arrays from their diagnostic samples had already been analysed as part of Chapters 4 and 5, including 7 BL patients and 2 DLBCL patients. Two patients were acquired from the Queen Elizabeth Central Hospital, Blantyre, Malawi. One was diagnosed with endemic BL and the other was considered to be a BCLU patient. The cohort is described in Figure 2.5

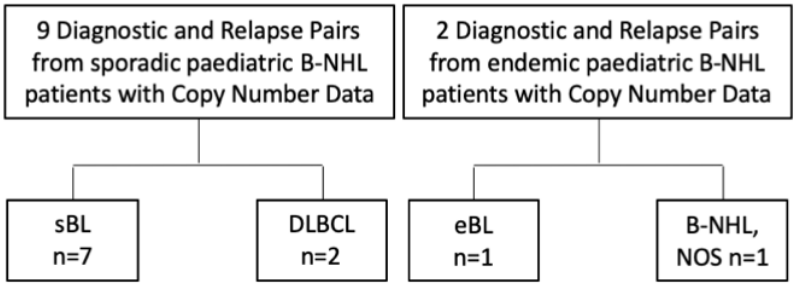


Figure 2.5 - Detail of the patients analysed in Chapter 7 of this thesis. 11 diagnostic and relapse pairs were investigated.

2.3.2 Pathology Review of patient samples acquired from the CCLG Biobank

As part of a larger study in the Rand Lab and with the support of Children Cancer and Leukaemia Group (CCLG) Tissue Bank, we established the largest cohort of clinically annotated paediatric B-NHL tumour samples taken from children diagnosed and treated in the UK. As expected from current survival data, 15% of children had relapsed/refractory disease, of which 9 have matched relapse samples. A pathology review was conducted to confirm that samples included in the study were from B-NHL patients and appropriate for use in the study, as well as confirming the diagnosis of the B-NHL subtype. Patients diagnosed with the following B-NHL subtypes were included in the analysis: Burkitt lymphoma (BL), diffuse large B cell lymphoma (DLBCL), Burkitt-like Lymphoma with 11q aberration (BLL-11q) and high grade B cell Lymphoma, not otherwise specified (HGBL, NOS). Patients that could not be confidently prescribed to a B-NHL subtype were considered B-NHL, NOS (not otherwise specified). Pathology review was performed on patient samples with available formalin-fixed paraffin-embedded (FFPE) material. Where available, unstained sections were cut from FFPE blocks and IHC (immunohistochemistry) stained for CD10, CD20, CD3, BCL2, BCL6 and Ki67 by the Cellular Pathology Department of the Royal Victoria Infirmary (RVI), Newcastle. Haematoxylin and eosin (HE) staining was also requested if fresh frozen (FF) material was available. The review of material was performed by three haematopathologists from the RVI: Drs Christopher Bacon, Despina Televantou and Katrina Wood and consensus diagnoses were reported. Confirmation of *MYC* status was performed when possible by fluorescence *in-situ* hybridisation (FISH). Cases with no consensus on diagnosis between three pathologists but confirmed to be B-NHL were reported as B-NHL, NOS (B cell non-Hodgkin lymphoma, not otherwise specified). For cases with no available FFPE material for pathology review, Dr Christopher Bacon performed a review of the original pathology report from the local treatment centre to ensure that the initial diagnosis was appropriate. Patients diagnosed prior to publication of the 2016 WHO Classification of Lymphoid Neoplasms were reviewed to determine whether the initial diagnosis was still appropriate, including new diagnoses of HGBL, NOS and BLL-11q.

2.3.3 Pathology Review of patient samples acquired from QECH, Malawi

A collaboration between Newcastle-upon Tyne Hospital Trust/RVI and the Queen Elizabeth Central Hospital in Blantyre, Malawi has allowed us to generate a cohort of 105 paediatric BL cases from the endemic subtype. Pathology review of the endemic BL cohort acquired from the Queen Elizabeth Central Hospital in Blantyre, Malawi had previously been performed within the group. Diagnoses were confirmed by pathology review according to the World Health Organisation criteria using immunocytochemistry for CD20, CD3, CD10, BCL2, BCL6 and Ki67, *in situ* hybridization for Epstein Barr Virus-encoded small RNA (EBER) and Fluorescence *in-situ* hybridisation (FISH) for the presence of a *MYC* translocation.

2.3.4 Fluorescence in-situ hybridisation (FISH) on unstained imprints from fresh frozen material

FISH was performed to determine the *MYC*-rearrangement status of patients with tumour material. Superfrost glass slides with tumour imprints that were received from the Cellular Pathology department of the RVI were marked using a diamond point marker to highlight the location of the imprint on the slide. Slides were marked from underneath so as not to damage the tumour imprint. Slides were already marked with patient samples and the date of set-up and probe name were written onto the slides in pencil. 2µl of FISH probe was pipetted onto the slide within the marked area. The area was covered with a 10mm diameter coverslip and sealed using a rubber cement. The slides were loaded onto a Thermobrite Hybridiser and the foam pads in the chamber were dampened to keep the environment humid. The Thermobrite Hybridiser was programmed to denature for 2 minutes at 80°C and then hybridise at 37°C overnight.

A water bath was prepared at 72°C before removing slides from the hybridiser the next morning. A plastic Coplin jar was filled with wash solution (0.4xSSC/0.003% NP-40) and left to heat to 72°C in the water bath. The coverslip was removed from the slide using needle point tweezers, removing any residue of rubber cement along with it. The slides were immediately submerged and washed in the 72°C wash solution for 2 minutes. Slides were then subsequently transferred directly to another Coplin jar of 2xSSC at room temperature for 2 minutes. Slides were removed from the 2xSSC

solution and 8µl of Vector Shield Mounting Medium with DAPI was applied to the slide and covered with a 22x50mm coverslip (0.13mm thickness). Slides were then viewed after 15 minutes or stored in a dark box at 4°C for scoring at a later date.

2.3.5 Fluorescence in-situ hybridisation (FISH) on unstained sections from FFPE material

The protocol for performing FISH analysis on samples with FFPE material required additional steps compared to the preparation of fresh frozen imprints. Slides were received from Cellular Pathology on Superfrost glass slides. Slides were submerged in a Coplin jar of Xylene for 10 minutes inside a fume hood. Xylene was then discarded and the wash was repeated with fresh Xylene. After the second Xylene wash, slides went through a rehydrating gradient of ethanol preparations. The slides were submerged in a Coplin jar of 100% ethanol for 1 minute at room temperature, followed by 85% ethanol for 1 minute, 70% ethanol for 1 minute and washed in distilled water for 1 minute. After the wash with water, another 1 minute wash with distilled water was performed.

200µl of Cytocell Tissue Pre-treatment kit reagent 1 was pipetted onto each slide and they were covered with a 22x50mm coverslip (0.13m thickness) and sealed with rubber cement. Slides were then placed on the Thermobrite Hybridiser at 99°C for 15 minutes. Once removed from the hybridiser, coverslips were removed with a pair of needle point tweezers and the slides were washed again in distilled water for 1 minute, twice. 2 drops of Cytocell Tissue Pre-treatment kit reagent 2 were applied to the slide and covered with 22x50mm coverslips (0.13mm thickness), sealed with rubber cement and incubated at room temperature for 30 minutes before being washed in distilled water for 1 minute, twice.

The slides were then dehydrated by being submerged in 70% ethanol for 1 minute, 85% ethanol for 1 minute and then 100% ethanol for 1 minute. Once removed from the Coplin jars, slides were left to air dry on paper towels. 4µl of FISH probe was pipetted onto the slide and covered with a 19mm diameter round coverslip, sealed again with rubber cement. The slides were then placed on the Thermobrite Hybridiser and the foam pads were soaked in water. The Thermobrite Hybridiser was programmed to denature at 75°C for 2 minutes and then incubate at 37°C overnight.

The following morning the coverslips were removed from the slide with needle point tweezers and the slides were incubated in a Coplin jar with 0.4xSSC/0.003% NP-40 for 2 minutes in a 72°C water bath. The slides were transferred to another Coplin jar of 2xSSC for 2 minutes at room temperature before 8µl of Vector Shield Mounting Medium with DAPI was dropped onto the slide and a 22x50mm (0.13mm thickness) coverslip was applied. The slides were left to incubate for 15 minutes before being visualised, or were stored at 4°C for later scoring.

2.3.6 FISH slide Scoring

FISH slides were scored on an Olympus BX61 fluorescence microscope. At least 50 cells were scored with the fluorescent patterns recorded in a lab book and tallied. Each slide was scored by a minimum of two trained researchers independently. A panel of normal bone marrow preparations hybridised with the same probe as the tumour sections/imprints were scored by each researcher. This was performed to calculate a minimum cut off of aberrant cells reported for each individual, which is required to ensure that results were consistent between researchers. Five normal bone marrow samples were scored and the mean + 3 standard deviations of abnormal patterns across all five marrow preparations was calculated as a cut-off. If the ratio of abnormal cells identified by an individual in a case was above cut-off then it was considered a positive case. Results were then compared between both scorers, and discrepancies were re-scored by both individuals.

2.3.7 Nucleic Acid Extraction from Fresh Frozen Tumour Samples

DNA and RNA were extracted from fresh frozen tumours using the Qiagen Allprep DNA/RNA/Protein mini kit (50 samples) at the beginning of the study. Later, additional samples were extracted instead with the Qiagen Allprep DNA/RNA/miRNA Universal kit according to the manufacturer's guidelines to achieve total RNA extraction from limited fresh frozen material. The DNA extraction method was identical in both protocols. RNA extractions were treated with the Ambion Turbo DNA-free DNAse kit to remove residual genomic DNA. This allowed the RNA aliquots to be used for RNA sequencing in the future. Total RNA was eluted into nucleic acid-free water and DNA aliquots were eluted into EB buffer.

2.3.8 Nucleic Acid Extraction from FFPE Tumour Samples

DNA and RNA were extracted from FFPE tumour material using the Qiagen DNA/RNA FFPE kit. Deparaffinisation was performed initially on FFPE curls cut by Cellular Pathology on a microtome. In 1.5ml microcentrifuge tubes curls were treated with 800µl of Xylene. Tubes were vortexed thoroughly and incubated for 5 minutes at room temperature. The tubes were then spun down in a centrifuge at 12,000G for 2 minutes. The supernatant was discarded by pipette. The deparaffinisation step was repeated once more per sample. After removing the supernatant for the second time, 800µl of 100% ethanol was added to the microcentrifuge tubes and vortexed thoroughly. Without incubation, the tubes were then spun down using a centrifuge for 2 minutes at 12,000G. The tubes were then left to incubate on a heat block set to 55°C with the caps off to allow residual liquid to evaporate off. From this point onwards the standard protocol for the Qiagen AllPrep DNA/RNA FFPE kit was followed.

2.3.9 DNA Quality Control and Quantification

DNA concentration was measured by Qubit 3.0 Fluorometer (ThermoFisher, USA). The Thermofisher Scientific Qubit dsDNA Broad Range Assay was used to calculate the concentration of double stranded DNA specifically. A new standard curve was created for every batch of 10 samples. RNA concentration and integrity (RIN score) were determined by Agilent Bioanalyzer 2100 using the Agilent RNA 6000 Nano kit and later the Agilent Tapestation 4200 using the Agilent RNA ScreenTape kit according to the manufacturer's protocols. DNA and RNA were stored at -80°C in 1.5ml microcentrifuge tubes.

2.3.10 Polymerase chain reaction (PCR) of patient DNA for Sanger sequencing

Patient DNA was diluted to a stock concentration of 10ng/µl. A minimum quantity of 50ng of patient DNA was added. A PCR master mix was made up using the AmpliTaq Gold DNA Polymerase kit as specified in Table 2.2.

Master Mix	**x1**
10X PCR Buffer	5µl
dNTPs	0.4µl
$MgCl_2$	3µl
AmpliTaq Gold Polymerase	0.25µl
5µM Forward primer	1µl
5µM Reverse primer	1µl
Nucleic acid-free water	Up to 45µl
TOTAL	45µl

Table 2.2 – Components of the PCR master mix for one reaction.

The AmpliTaq Gold Polymerase was taken out of the freezer as the last step to preserve the quality of the polymerase. 45µl of master mix was aliquoted to each well of a 200µl 96-well skirtless PCR plate per sample with a water control. 5µl of patient DNA (at 10ng/µl) was added to each plate well, as well as 1µl of 5µM Forward primer and 1µl of 5µM Reverse primer. 5µl of nuclease-free water was added instead of DNA to the water control well.

The plate was sealed with strip tube lids before being vortexed and spun down briefly to ensure the reaction volume was well mixed by pipetting. The plate was run on an Agilent SureCycler 8800 thermal cycler. A pre-made PCR program was selected as specified in Table 2.3.

Step	Temperature	Time
Initial Denaturation	95 °C	10 minutes
35 Cycles	95 °C	20 seconds
	54 °C	1 minute
	72 °C	1 minute
Final Extension	72 °C	5 minutes
Hold	10 °C	Indefinitely

Table 2.3 - PCR machine program for the amplification of patient DNA.

2.3.11 Gel electrophoresis of PCR products

The presence of successful PCR product amplification was confirmed by gel electrophoresis. A stock of 50x TAE buffer was made up as shown in Table 2.4.

Reagent	Quantity
Tris Base	242g (dissolved in 750ml of water)
Glacial acetic acid	57.1ml
0.5M EDTA (adjusted to pH 8.0)	100ml
Nuclease-free water (top up)	Adjust to final volume of 1L

Table 2.4 - Protocol for making up 1000ml of 50X TAE buffer for DNA gel electrophoresis.

Analytical grade agarose powder was weighed out appropriately to make up a 1.2% agarose gel. 0.84g of electrophoresis-grade agarose powder was added to 70ml of 1X TAE buffer in a conical flask. After the powder was dissolved via microwave heating in short bursts for 2.5 minutes, 7µl of GelRed (10,000x concentrated) was added to the mixed 1.2% agarose solution and the mixture was poured into a pre-assembled DNA gel tank with 16-well combs in place inside. The agarose gel was left to set for approximately 20 minutes, making the sure gel was completely set. 1X TAE buffer was poured over the gel until the electrodes in the tank were sufficiently submerged. The combs were removed after the 1X TAE had been added. 4µl of PCR product per sample was combined with 1µl of DNA loading dye on a sheet of parafilm and then carefully pipetted into each well on the gel. 2µl of DNA Hyperladder (100bp) was added to the first and last well of each row. Once the lid was placed on the tank it was connected to a BioRad powerpack and run for 40 minutes at 70 volts. The gel was then visualised under ultraviolet light using a BioRad ChemiDoc machine. The size of the bands on the gel were compared to the predicted product size to ensure the right region of genome was amplified. Negative water controls were checked to ensure nothing had been amplified in them, and patient samples were checked to ensure that only one band had been amplified.

2.3.12 Purification of PCR products

PCR products in 96-well plates were purified by Eurofins Genomics, Germany. Purification of products that were amplified in 8-well strip tubes was performed in-house using the QIAquick PCR Purification kit as per the manufacturer's protocol. Purified products were aliquoted and sent to Eurofins Genomics, Germany for Sanger sequencing.

2.3.13 Analysis of Sanger sequencing data

When Sanger sequencing results were received the data was visualised using FinchTV software. In FinchTV, Sanger chromatogram files (.ab1 file format) were checked manually to determine whether any nucleotides exhibited two peaks on top of each other, indicating that more than one nucleotide was detected at that locus. This would represent a heterozygous mutation or a mutation that was present in a subclone of the patient sample (either heterozygous or homozygous). DNA sequences generated from the Sanger chromatogram were next screened using BLAT against sequences on Ensembl to detect mutations that were homozygous (Cunningham *et al.*, 2019).

2.4 Bioinformatics Methods

2.4.1 Copy Number Data Processing of Affymetrix SNP 6.0 array

25 paediatric B-NHL fresh frozen samples provided by the Children's Cancer and Leukaemia Group (CCLG) were investigated using a high-resolution SNP microarray platform. Tumour DNA was extracted using the Qiagen All Prep DNA/RNA/Protein extraction kit and sent to AROS (Applied Biosystems, Denmark) to be hybridised to an Affymetrix Genome Wide SNP 6.0 Array. The raw cel probe intensity files were analysed using the SNP Rank segmentation algorithm and manually annotated to exclude obvious artefacts. Five cases were not re-analysed using Affymetrix CytoScan HD or OncoScan v2 MIP arrays for analysis in Chapters 4 onwards and the original SNP6.0 data was used for copy number analysis.

2.4.2 Copy Number Data Processing of Affymetrix CytoScan HD array

DNA from 79 diagnostic samples of B-NHL patients acquired through the CCLG biobank were sent to AROS (Applied Biosystems, Denmark) to generate copy number data using the Affymetrix CytoScan HD array. Raw cel intensity files were analysed using Nexus Copy Number 10.0 with the SNPRank segmentation algorithm. Thresholds for copy number gain and loss were set to 0.15 and -0.15 respectively.

2.4.3 Copy Number Data Processing of Affymetrix OncoScan 2.0 MIP array

DNA from 26 diagnostic samples of B-NHL patients acquired through the CCLG biobank and 105 diagnostic samples from the endemic BL Malawi cohort were sent to AROS (Applied Biosystems, Denmark) to be analysed on the Affymetrix OncoScan 2.0 molecular inversion probe array. Raw pairs of cel intensity files were analysed using Affymetrix OncoScan Console version 1.3 to generate OSCHP files with normalised log2 ratio data and segmentation data produced via the proprietary TuScan algorithm using the default settings. Due to the nature of the algorithm and the fact that it calculates an estimate of the proportion of aberrant cells in a sample, hard coded thresholds for gain and loss calling are not possible. Segmented copy number calls generated by the TuScan algorithm were viewed and manually checked in Nexus Copy Number 10.0. The TuScan algorithm was favoured due to its reliability when calling subclonal events or events in samples with low tumour content. This was particularly important when working with fine needle aspirations taken from patients in Malawi when the tumour content was expected to be lower.

2.4.4 Copy Number Data Processing of External Data from GSE57612 and GSE21597

Both studies were performed on subsets of the same cohort dataset looking at BL (GSE21597) and DLBCL (GSE57612). Data was processed from raw CEL files using Nexus Copy Number 10.0 using the SNP Rank segmentation algorithm. Both Affymetrix GeneChip Human Mapping 250K Nsp array and Affymetrix GeneChip Human Mapping 250K Sty array data was available for 30/39 cases within GSE21597, which do not share probe IDs. In these 30 cases the data was merged using Nexus Copy Number 10.0 to give a final resolution of 500K. A systematic correction file for the Affymetrix GeneChip Human Mapping 500K array was applied to both data types. Due to the noisy nature of the data the segmentation settings were changed to decrease the threshold for copy number gain from 0.15 to 0.1 and the significance threshold was set to 5.0×10^{-9}. Clinical data, including patient age, follow up and relapse status was made available by the authors (Scholtysik *et al.*, 2010; Scholtysik *et al.*, 2015) and additional clinical information for both datasets was kindly provided by the authors on request.

2.4.5 Copy Number Data Processing of External Data from GSE19459

The dataset is associated with a study into noise reduction on aCGH arrays (Lepretre *et al.*, 2010). 20 DLBCL patient samples were hybridised on an IRCL Human Genome CGH Microarray 44B G4410B against Promega control DNA. Data was processed from raw text files using Nexus Copy Number 10.0 using FASST2 segmentation. Segmentation thresholds for gain and loss were increased to 0.2 and -0.23 respectively to reflect the noise level of the samples from this study.

2.4.6 Copy Number Data Processing of External Data from GSE11318

The dataset was associated with a study investigating copy number abnormalities in different DLBCL subtypes (Lenz *et al.*, 2008). 203 DLBCL patients were analysed using an NCI NimbleGen Homo sapiens HG17 Whole Genome 385K Tiling Set (version 1). Of the 203 samples published, 27 were unclassifiable as to their DLBCL subtype, and 31 were primary mediastinal DLBCL so were excluded from analysis. Segmentation thresholds for gain and loss were increased slightly to 0.17 and -0.20 respectively.

2.4.7 Copy Number Data Processing of External Data from GSE12906

The dataset was published alongside a Letter to Nature by Kato *et al*. (Kato *et al.*, 2009). 62 DLBCL samples were used in our investigation. The platforms used in the studies were a combination of Affymetrix SNP arrays; Affymetrix GeneChip Human Mapping 250K Nsp arrays and Affymetrix GeneChip Human Mapping 250K Sty arrays and the lower resolution Affymetrix GeneChip Mapping 50K XbaI arrays and Affymetrix GeneChip Mapping 50K HindIII arrays. Only site of disease and diagnosis by histology was made available. Data analysis was performed from raw .cel files using the SNP-FASST2 algorithm. Segmentation thresholds for gain was increased slightly to 0.1.

2.4.8 Copy Number Data Processing of External Data from GSE56884

The dataset comprised of 81 tumour samples from 21 follicular lymphoma patients, 31 transformed DLBCL patients and 29 *de novo* DLBCL patients. Only *de novo* DLBCL patients were included in our investigation. The data was published to accompany Kwiecinska *et al.* (2014) and was accessed from GEO. Extensive clinical data was published including patient age at diagnosis, gender, DLBCL subtype,

tumour stage, outcome, follow-up time, relapse status and treatment strategy. Diagnosis was made by immunohistochemistry according to WHO 2008 classification. Processed log ratio data was made available in the form of tab-delimited text files. The array used was a custom-made CGH array with an approximate resolution of 1Mb manufactured by the Wellcome Trust Sanger Institute, Hinxton, UK. The processed data was visualised in Nexus Copy Number 10.0 using 0.2 and -0.2 as gain and loss segmentation thresholds respectively.

2.4.9 Copy Number Data Processing of External Data from GSE54303

The dataset involved 332 untreated lymphoma patients. Only 117 DLBCL and 24 BL patients were used in this investigation, totalling 141 samples. The dataset was published as part of a publication by Suguro *et al*. (2014) in *Cancer Science*. This study investigated the prognostic role of clonal heterogeneity in lymphoma using a custom-made BAC array with approximately 2000 probes. Raw log ratio data was available via GEO and clinical information was available for survival status, follow up in years and IPI score. No information about patient age was made available. Patients were diagnosed at the Aichi Cancer Center Hospital and Nagoya University Hospital, Japan according to the WHO classification. Thresholds for segmentation were set at 0.15 and -0.15 for gain and loss respectively.

2.4.10 Copy Number Data Processing of External Data from Havelange et al. 2016

Although copy number array files were not available for the samples investigated in this publication, the authors of the study published the start and end positions of regions of genomic aberration in the supplementary information, and these were used to create segmented data files that could be loaded into Nexus Copy Number 10.0 and analysed alongside data from other studies. The study performed copy number detection using Affymetrix Mapping 100K-XbaI arrays.

2.4.11 Inclusion/Exclusion criteria

Samples from in-house and external datasets were chosen based on available diagnostic information. This investigation focused only on Burkitt lymphoma, diffuse

large B cell lymphoma, Burkitt-like lymphoma with 11q abnormality and cases which are known to be B-NHL but where the subtype was not determined. Patients presenting with post-transplant disease, lymphoblastic lymphoma, T cell (or T cell-rich) lymphoma, mediastinal lymphoma, pre-B cell lymphoma, unspecified lymphoma and those with no tumour sample available were excluded from our analysis.

2.4.12 Data Analysis with Nexus Copy Number 10.0: Segmentation algorithms

Several factors determined the analysis and segmentation parameters used in our investigation. Two segmentation algorithms included in Nexus were used: Rank segmentation and fast adaptive states segmentation technique (FASST2). Both algorithms use log-ratio values and genomic positioning of probes to segment data into regions of statistically significant copy number change. Rank segmentation is based on the circular binary segmentation (CBS) algorithm but is significantly faster to run than standard CBS. FASST2 is a Hidden Markov Model (HMM) based algorithm and uses this method to estimate segment values between expected copy number states, and further processing this information to assign these values to events. Both segmentation algorithms incorporated into Nexus Copy Number 10.0 also exist as a “SNP” version, designed specifically for copy number arrays with allelic frequency data for SNPs associated. The algorithm works in fundamentally the same way in each case, but the SNP form uses allelic information to complement log-ratio values in copy number detection and ploidy calculation for the sample.

2.4.13 GISTIC1.0 analysis to detect driver aberrations

Developed by the Broad Institute, Genomic Identification of Significant Targets in Cancer (GISTIC) is a tool designed to identify regions within copy number data that are significantly aberrant. The first iteration of GISTIC is implemented into Nexus Copy Number 10.0 and calls peaks at locations of the genome with the most frequent aberrations across a group of samples. Both frequency of the abnormality and the magnitude of the copy number change are taken into account by the algorithm to produce a G-score. The software identifies a tight peak using this G-score, and a wide peak using a leave-one-out method to ensure the tight peak is not defined purely based on one sample (Beroukhim *et al.*, 2007). The functionality of GISTIC relies on the assumption that driver events caused by copy number variation occur

more frequently than passenger events in a cohort of patient samples. Therefore, the most significant regions identified by GISTIC should theoretically be disease drivers.

2.4.14 GISTIC2.0 analysis to detect driver aberrations

As successor to GISTIC1.0, GISTIC2.0 was developed to more reliably detect significant targets in cancer from somatic copy number data. The second iteration of GISTIC2.0 splits genomic events into focal and arm-length events to further characterise genomic events (Mermel *et al.*, 2011) (Figure 2.6).

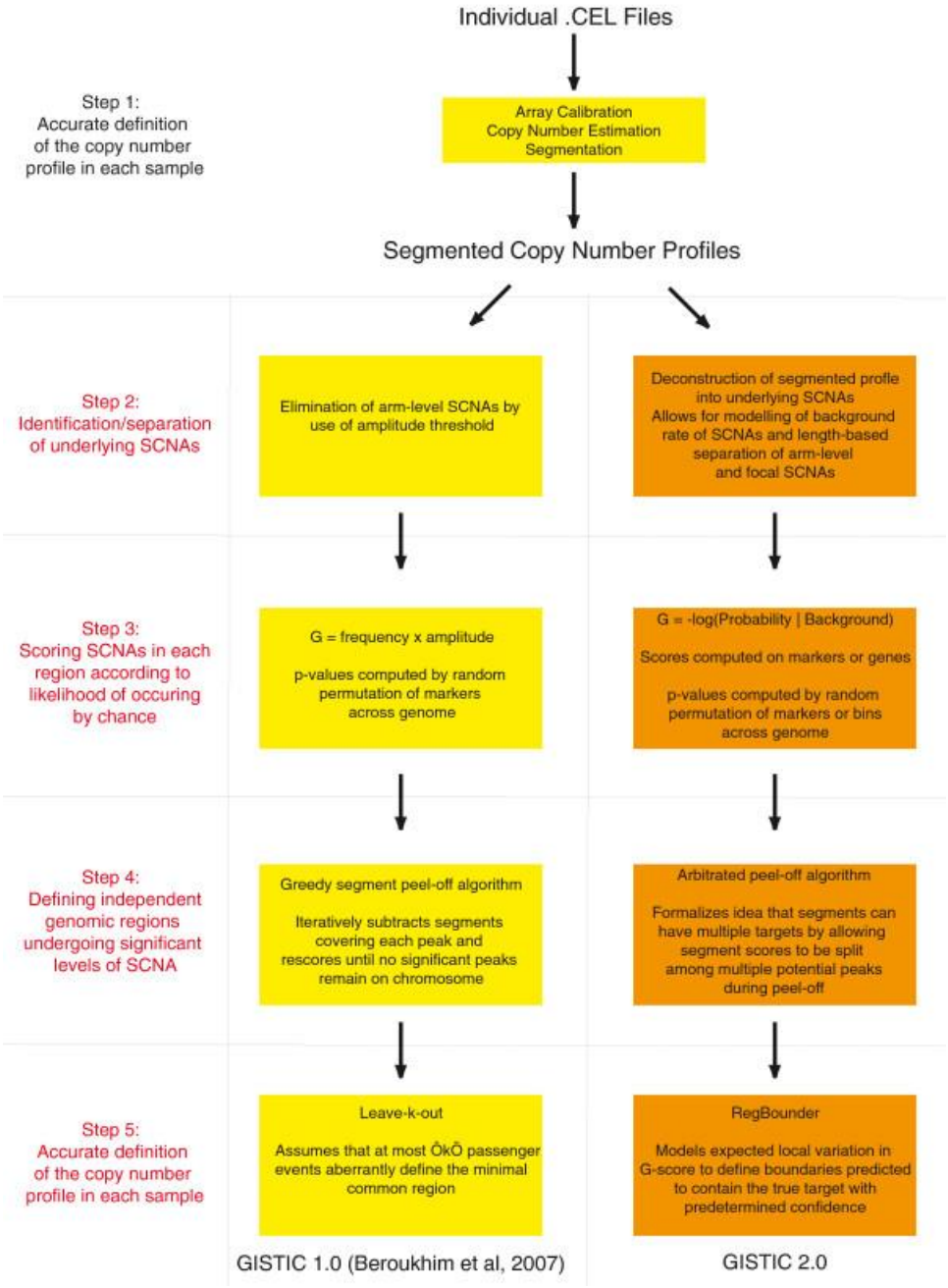


Figure 2.6 - A flow diagram describing the differences between GISTIC1.0 and GISTIC2.0. Taken from Mermel et al.(2011).

While GISTIC1.0 generates a G score for recurrent copy number aberrations calculated from the amplitude of the log ratio multiplied by the frequency of the aberration within the cohort, GISTIC2.0 uses a more advanced method. The updated algorithm used by GISTIC2.0 computes a G score based on the negative log probability of the region being part of the background pattern. Depending on the settings provided to GISTIC2.0 on the command line, this can be either on a gene basis or a marker basis. As markers are surrogates for probes, and the cohort being analysed is comprised of multiple array platforms with different probe sets, it was decided that gene-based scoring was more appropriate. A function new to GISTIC2.0 then allows the tool to distinguish between multiple peaks/targets within a segmented region (i.e. a chromosome arm) using a peel-off function. This functionality is most useful when there are multiple recurrent peaks on a single chromosome that may be covered by the same copy number event in some samples.

Another improvement on the previous algorithm allows the minimal common region of gain or loss to be more appropriately defined using a probabilistic method. This result gives both a wide peak, which is the more conservative estimate of the true bounds of the region of interest, while the narrow peak region is the genomic region that GISTIC2.0 is confident to be focal to the true aberration.

GISTIC2.0 requires segmented copy number data with log ratio values for the entire genome, even if no events are called. Regions with missing data in any sample are excluded from the entirety of the cohort analysis. Centromeric regions were covered inconsistently by different array platforms so this ensured that events around the centromere were called appropriately regardless of the degree of coverage.

GISTIC2.0 analysis was performed as part of Chapter 4 and Chapter 6, while the original iteration was utilised in Chapter 3. Parameters for each separate GISTIC2.0 analysis are available Appendix A and in the shell scripts run by the docker container in the GitHub repository associated with this thesis (https://github.com/anewman66/phd-thesis-2020). The dockerfile used to generate the docker container was acquired from GitHub (https://github.com/sandertan/docker-gistic2) and the version of GISTIC2.0 was GISTIC2.0.23. The container was run with a local file mounted to allow uploading of data to the container without stopping and

resetting the container each time an edit was made to a script. Plotting GISTIC2.0 focal regions against G score was performed using maftools in R for analysis in Chapter 4.

2.4.15 Whole exome sequencing analysis

Whole exome sequencing data was analysed to screen the protein coding regions of the cancer genome for mutations. Wherever possible both the tumour and matched normal DNA were sequenced. In total, whole exome sequencing data was available for 90 B-NHL cases from the CCLG cohort. Of these, matched normal DNA was sequenced in 41 cases and tumour alone was sequenced for 49. A number of whole exome sequencing capture kits were utilised in this study and library prep was performed either at AROS (Applied Biosystems, Denmark) or the Genomics Core Facility (Centre for Life, Newcastle, UK). Kits used were Agilent SureSelect Human Exome v4 kit, Illumina Nextera DNA Exome and the Twist Human Core Exome kit. Sequencing was performed on Illumina HiSeq 4000 at AROS and either the Illumina MiSeq instrument or the Illumina NovaSeq 6000 at the Genomics Core Facility. All sequencers generated paired end reads of 100bp.

The analysis pipeline involved quality control (QC) of the raw reads, alignment to a reference genome, aligned read processing, variant calling and functional annotation of variants. Fastq files containing raw reads from the sequencer were the starting point for the analysis. Quality control was performed by applying the FastQC tool to each fastq file. The most important metric to check at this stage was that the base quality score was not associated with a particular base position on the read. Even high quality Illumina sequencing data often exhibits lower base quality scores toward the end of the read, but this can be more pronounced in poor quality sequencing (Andrews, 2010; Opitz *et al.*, 2019). After checking the per base sequence quality using FastQC it was sufficiently high across all samples so no cases were excluded.

Bioinformatics processing and analysis of sequencing data was performed using the FMS Cluster at Newcastle University. The cluster uses the Son of Grid engine (SGE) as a job scheduler and all scripts were written to be scheduled by SGE. Alignment, processing and variant calling were all performed using a pipeline written by Dr Matt Bashton at Newcastle University. The scripts are available from Dr Bashton's GitHub

page (https://github.com/MattBashton/MB-GATK-SGE). Scripts were written in the bash language and process NGS data using GATK3.8. The pipeline took pairs of raw fastq files and aligned the reads using the Burrows-Wheeler aligner, maximal exact matches (BWA-MEM). BWA-MEM is an alignment tool based on the Burrows-Wheeler transform, and the -MEM iteration of the algorithm is designed specifically for aligning Illumina paired end sequencing data with reads over 70bp, which was appropriate for our data. Reads were aligned to the hg19 1000 Genome Project Decoy genome (b37). BWA-MEM outputted sequencing alignment mapping (SAM) files which were converted to binary alignment mapping (BAM) files and were sorted and indexed. At this point the aligned reads could be viewed in Integrative Genomics Viewer (IGV) developed by the Broad Institute. Further processing was required before variant calling could be performed.

Duplicate reads were marked using Picard, but not removed. Marked duplicates were retained as they can still be helpful. Marking duplicate reads is important to ensure that errors are not overrepresented when calling variants. The next step was to perform realignment target creation, where indels (insertions and deletions) are highlighted as targets for indel realignment. As somatic indels are not present in the reference genome, realignment must be performed to ensure that the reads surrounding these indels are correctly mapped and that the indels are not artefacts (Li and Durbin, 2009). Base Q score recalibration (BQSR) was performed as there are inherent biases in the way that Illumina sequencers assign quality scores. Adjustment of these scores is required as inaccurate scores are known to have an effect on variant caller results (Tian *et al.*, 2016). Reads were printed with recalibrated quality scores to new BAM files for further analysis.

The next step involved variant calling with MuTect for tumour only data and Mutect2 for cases with matched tumour and normal sequencing data.

The original MuTect tool was utilised for tumour only analysis. The tool detects variants using four primary steps (Cibulskis *et al.*, 2013). Step one is to remove low quality sequencing data by filtering reads based on the criteria listed in Table 2.5

Read Quality Filters for MuTect
(a) Mapping Quality score > 0
(b) Base quality score ≥ 5
(c) If there is an overlapping read pair, and both reads agree the read with the highest quality score is retained otherwise both are discarded.
(d) Sum of the quality scores of the mismatches ≤ 100
(e) < 30% of bases have been soft-clipped
(f) Reads that have not been mapped by "mate rescue" of BWA1,2 (BAM XT tag ≠ "M")

Table 2.5 - A table listing quality control filters for input reads to MuTect. Adapted from Cibulskis et al., (2013).

Step two is to detect variants using a Bayesian classifier. This function compares two models for each locus: a reference model where there is no variant at the locus and any calls are due to sequencing errors, and a variant model where a true variant is found at the locus. An allelic fraction value is calculated to determine the proportion of reads that support the variant model. This value is used instead of assuming heterozygosity so that the tool is more sensitive in cases where tumour purity is lower or subclonal events exist (Cibulskis *et al.*, 2013). The variant is considered a candidate true variant if the log odds value exceeds a predetermined value based on other criteria, namely a predefined false discovery rate and the expected frequency of mutations.

Step three involves filtering called variants based on known parameters associated with false positive calls. Six filters are implemented, as well as comparison against a panel of normals from the same demographic as the cohort of patient samples (Table 2.6).

Filter Name	**Class**	**Description and Default Thresholds**
Proximal Gap	HC	Remove false positives caused by nearby misaligned small insertion and deletion events. Reject candidate site if there are ≥ 3 reads with insertions within an 11-bp window centered on the candidate mutation, or if there are ≥ 3 reads with deletions within the same 11-bp window
Poor Mapping	HC	Remove false positives caused by sequence similarity in the genome, leading to misplacement of reads. Two tests are used to identify such sites: (i) Candidates are rejected if ≥ 50% of the reads in the tumor and normal have a mapping quality of zero (although mapping quality zero reads are discarded in the short-read preprocessing this filter reconsiders those discarded reads); (ii) Candidates are rejected if they do not have at least a single observation of the mutant allele with a confident mapping (i.e. mapping quality score ≥ 20).
Triallelic Site	HC	Reject false positives caused by calling tri-allelic sites where the normal is heterozygous with alleles A/B and MuTect is considering an alternate allele C. Although this is biologically possible, and remains an area for future improvement in mutation detection, calling at these sites generates many

		false positives and therefore they are currently filtered out by default. However, it may be desirable to review mutations failing only this filter for biological relevance and orthogonal validation and further study the underlying reasons for these false positives.
Strand Bias	HC	Reject false positives caused by context specific sequencing errors where the vast majority of the alternate alleles are observed in a single direction of reads. We perform this test by stratifying the reads by direction and then applying the core detection statistic on the two datasets. We also calculate the sensitivity to have passed the threshold given the data. Candidates are rejected when the strand specific LOD is < 2.0 in directions where the sensitivity to have passed that threshold is ≥ 90%.
Clustered Position	HC	Reject false positives caused by misalignments hallmarked by the alternate alleles being clustered at a consistent distance from the start or end of the read alignment. We calculate the median and median absolute deviation of the distance from both the start and end of the read and reject sites that have a median ≤ 10 (near the start/end of the alignment) and a median absolute deviation ≤ 3 (clustered)
Observed in Control	HC	Eliminate false positives in the tumor by looking at the control data (typically from the matched normal) for evidence of the alternate allele beyond what is expected from random sequencing error. A candidate is rejected if, in the control data, there are (i) ≥2 observations of the alternate allele, or they represent ≥ 3% of the reads; and (ii) their sum of quality scores is > 20.
Panel of Normals	HC+PON	Reject artifacts and germline variants by inspecting a panel of normal samples and rejecting candidates that are present in two or more normal samples

Table 2.6 - A table describing variant filters utilised in the high confidence method of MuTect analysis.

Step four involves determination of somatic and germline variants based by another Bayesian classifier. All variants were designated as either “somatic”, “germline” or “variant”. The variant call is used for variants that are confirmed to be present in the tumour but where the status in the matched normal is undeterminable. This call was not required in our cohort as all samples with matched normals were analysed via Mutect2.

Mutect2 is incorporated in GATK3.8 and is run on processed BAM files to produce a list of variants (Benjamin, 2019). Sharing a lot of code with the existing Haplotype Caller (HC), Mutect2 identifies active regions where somatic variation is likely to occur by comparing tumour and normal log odds. If the log odds value for a variant exceed the threshold set in Mutect2 then the intervals around the active site are considered for realignment. A series of haplotypes are generated and ranked based on their score. A pair-Hidden Markov Model (HMM) function for pairwise sequence alignment assesses the likelihood of each read to have been sequenced against each candidate haplotype for both the tumour and normal samples. A likelihood ratio for each allele is generated using this tool and potential variants are filtered based on calls made by Mutect2. The risk of variants being present in the germline is

determined using TLOD (tumour log odds) and NLOD (normal log odds) values. If a variant NLOD was above 2.2 and TLOD was above 6.3 then the variant was not considered to be a germline risk. The NLOD threshold was set to a higher value of 5.5 if the variant was present in the COSMIC database of somatic variants in cancer (Tate *et al.*, 2019). NLOD is the likelihood that a variant is not present in the normal sample from the patient and the TLOD is the likelihood that the variant is present in the tumour sample. A panel of normal was not utilised in this analysis but the option exists within Mutect2.

Once variant calls were made using MuTect and Mutect2 functional annotation was performed to identify mutated genes and the predicted impact of variants on the gene function. This was performed using Variant Effect Predictor (VEP) from Ensembl with the "—everything" option added. The bash function "grep" was used to pull lines from the VEP output files of interest. In Chapter 4 this was done to identify variants in a given genomic region. In Chapter 5 this was performed to create one file containing all variants in the cohort for a given list of genes of interest. An R script was used to manually filter results for coding mutations, recurrent variants and for specific predicted impacts using tools such as PolyPhen and SIFT (Adzhubei *et al.*, 2013).

2.4.16 Statistical analysis

Chi-squared and Fisher Exact tests to compare the incidence of events in patient subgroups were performed using the base package of R. All R scripts used in this thesis are included in the associated GitHub repository (https://github.com/anewman66/phd-thesis-2020). Data for comparisons was either loaded into R from a text file or manually edited into the script for smaller comparisons. Comparisons of continuous variables within the cohort was performed using a student t-test in R.

Survival analysis was performed in R. A tab-delimited text file was read into R containing the sample name, follow-up data (time to event/last follow up) for appropriate survival parameters (Time to Progression, Risk of Relapse or Overall Survival) and covariates to test. For analyses where subsets of the cohort were analysed separately a column specifying inclusion and exclusion criteria was added for filtering via R. Univariate survival analysis was performed using the Cox Proportional Hazard method. Cox Proportional Hazard modelling makes the

assumption that the effect that a covariate has on the hazard ratio of a cohort is constant across time (Singh and Mukhopadhyay, 2011). This assumption test was performed in R to ensure that this assumption held true in each cohort analysed in this study.

Covariates with univariate p values below 0.1 were included in multivariate modelling by a forward selection method of statistical regression. The most significant covariate from univariate analysis (i.e. with the lowest p value) was included in the first model. The second model was generated from the first model alongside one of the other covariates. All possible second models were compared to the first model using a likelihood ratio test to determine if the addition of the second covariate had improved upon the first model. If more than one model significantly improved upon the previous, then the model with the most significant p value by likelihood ratio test was taken forward to the next stage of the forward selection. This process continued until no models significantly improved upon the last model by likelihood ratio test.

The Kaplan-Meier survival estimate function was performed to visualise the difference in relative risk between two patient subsets. Kaplan-Meier plots were generated in R using the survminer package. Input data for Kaplan-Meier plots was the same as used for Cox Proportional Hazard survival analysis. Scripts for producing Kaplan-Meier plots are saved to the GitHub repository associated with this thesis (https://github.com/anewman66/phd-thesis-2020).

2.4.17 Data visualisation

Copy number data figures were generated using Nexus Copy Number 10.0, R and IGV. Aggregate plots were generated in Nexus Copy Number 10.0 and reformatted using Inkscape. Cohort heatmaps presented in Chapter 4 and Chapter 6 were generated in IGV using a seg file generated from Nexus Copy Number 10.0. The image was saved as a vector-based image in .svg format and then reformatted and annotated in Inkscape. GISTIC2.0 plots were generated using the “maftools” R package.

Figures presenting whole exome sequencing reads were generated using IGV and reformatted and annotated using Inkscape. Lolliplots presented in Chapter 5 and

Chapter 6 displaying *TP53* mutations were generated in R using the "GenVisR" package. Scripts used to generate plots are available at the GitHub repository associated with this thesis (https://github.com/anewman66/phd-thesis-2020). Circos plots presented in Chapter 4 were generated using the Circos tool in the perl language by Dr Masood Zaka. The input files were aggregate plot files generated from Nexus Copy Number 10.0.

Chapter 3: Pilot Project and Meta-analysis of the Literature Data

3.1 Introduction

Relapse is a significant clinical challenge in treating paediatric B-NHL, with salvage rates as low as 17% (Fujita *et al.*, 2008). While the introduction of rituximab into frontline therapy has reduced rates of relapsed/refractory disease significantly (Minard-Colin *et al.*, 2016), treatment is still highly intense and toxic to the patient (Anoop *et al.*, 2012). Until salvage therapy is improved to be consistently successful then de-escalation of high-dose chemotherapy is not going to be possible. Currently there are no markers that can predict disease progression used in the clinic, despite previous studies identifying gain of 7q, deletion of 13q and amplification of 13q31.3 as putative prognostic markers (Poirel *et al.*, 2009; Nelson *et al.*, 2010; Schiffman *et al.*, 2011). The reason for these prognostic markers failing to reach the bedside is unclear, but the clinical need for translatable prognostic markers is still strong. It is possible that validation of prognostic markers could not be performed due to the low number of patients for phase II clinical trial recruitment. 7q gain and 13q loss were identified in the context of a large clinical trial cohort with extensive clinical data involving 238 cases (Poirel *et al.*, 2009) and validated in an additional cohort of 28 patients by FISH analysis (Nelson *et al.*, 2010). In contrast, 13q31.3 amplification was identified in a small cohort of 28 patients (Schiffman *et al.*, 2011) and has not yet been validated as a prognostic marker in a larger cohort. A recent study has suggested that 13q31.3 amplification is associated with paediatric disease and is significantly less frequent in the adult disease (Havelange *et al.*, 2016b). The study did not investigate the prognostic value of the abnormality, but the suggestion of *MIR17HG* gain being associated with paediatric disease is potentially incompatible with it being prognostic, as the childhood cases do better than adults.

Age is an important factor in the outcome of B-NHL, with adult patients doing significantly worse when compared to their paediatric counterparts (Pfreundschuh *et al.*, 2006; Poirel *et al.*, 2009). A significant factor driving this will be the use of different treatment protocols, and when treated uniformly in the past the outcomes in both age groups were comparable (Magrath *et al.*, 1996). However, important differences in disease biology between adult and paediatric B-NHL also exist (Sandlund and Martin, 2016). Treatment protocols in adults differ to those given in the paediatric setting, opting instead for a CODOX-M-IVAC regimen. The

FAB/LMB96-based protocols given to children are not as well tolerated by adult patients, and this is also an issue for older, more frail adult patients who cannot safely be treated on CODOX-M-IVAC and are generally given less intensive protocols such as dose adjusted EPOCH-R (Zhu *et al.*, 2018).

In the paediatric setting, B-NHL is a relatively rare haematological malignancy and studies investigating the genomic characteristics of the disease are hampered by small sample sizes (Poirel *et al.*, 2009; Scholtysik *et al.*, 2010; Schiffman *et al.*, 2011; Salaverria *et al.*, 2014; Havelange *et al.*, 2016b). These smaller studies often publish their raw data allowing for re-analysis of their cases, as well as larger meta-analysis of all previously published cases. At the start of this project, relatively few of our in-house cases had copy number array data or were clinically annotated for powerful survival analysis. To address this, data from our in-house cohort was combined with those of cases from the literature to perform a pilot study.

3.2 Chapter Aims

The specific aims of the work detailed in this chapter were as follows:

- To identify copy number data from the literature and optimise analysis of data from different platforms together.
- To process in-house copy number array data and evaluate bioinformatic tools to analyse results.
- To perform a preliminary investigation into recurrent copy number abnormalities in B-NHL in a combined cohort and identify aberrations associated with age and subtype.
- Investigate the prognostic value of previously-published markers associated with poor outcome.

3.3 Results

3.3.1 Literature search for B-NHL copy number data

A data search was performed using online data repositories Gene Expression Omnibus (GEO), ArrayExpress and the Database of Genotypes and Phenotypes

(dbGaP). The search was performed to find datasets that were generated on copy number platforms, including copy number/SNP arrays or aCGH (array comparative genome hybridisation) microarray datasets with raw or processed data available to download. Datasets had to include tumour material with a diagnosis of *de novo* DLBCL or BL. Nine studies were initially identified, but two were excluded before analysis due to the nature of the investigation. Of these, one study identified was excluded due to only constitutional germline array data being made available (GSE58718 (Conde *et al.*, 2014)) while another study was excluded as it only investigated post-transplant DLBCL diagnoses, which are known to differ genetically to *de novo* DLBCL (GSE69440).

In total, seven studies were identified which met the inclusion criteria (Table 3.1). Additional copy number data was obtained from a paper published in the British Journal of Haematology in 2016, after the search was performed (Havelange *et al.*, 2016b), totalling eight copy number datasets.

Accession	Disease type	Citation	Samples	Platform
GSE19459	DLBCL	Lepretre et al. 2009	20	IRCL Human Genome CGH Microarray 44B G4410B (probe version)
GSE11318	DLBCL	Lenz et al. 2008	145	NCI NimbleGen Homo sapiens HG17 Whole Genome 385K Tiling Set
GSE57612	DLBCL	Scholtysik et al. 2015	148	Affymetrix Mapping 250K Sty2 SNP Array
GSE12906	DLBCL	Kato et al. 2009	62	Affymetrix Mapping 250K Sty2 SNP Array, 250K Nsp, 50K Hind240, 50K Xba240
GSE21597	BL	Scholtysik et al. 2010	39	Affymetrix 250K Nsp and 250K Sty
GSE56884	DLBCL	Ichimura et al. 2014	29	Array CGH (Custom) 1Mb resolution
GSE54303	BL, DLBCL	Suguro et al. 2014	142	Array CGH (Custom) ACC human 2K ver6.0.
N/A	BL	Havelange et al. 2016	24	Affymetrix Mapping 100K

Table 3.1 - Summary of genomic datasets used in this study.

3.3.2 Patient Cohort

At the beginning of the project copy number data for 634 samples was collected from existing in-house and external datasets. A total of 25 diagnostic patient samples were included in the analysis (12 BL, 4 DLBCL and 9 B-NHL, NOS) and 609 were acquired from the literature (88 BL and 521 DLBCL). Age data was available for 410 samples allowing for age stratification. Table 3.2 details the breakdown of patients with age information by diagnosis. 18 years of age was used as the cut off between paediatric and adult cases. Outcome data was available for 36 paediatric BL cases.

	Burkitt Lymphoma	Diffuse Large B Cell Lymphoma	B-NHL, NOS	Total
Paediatric	49	19	9	77
Adult	26	316	0	342
Unknown	25	190	0	215
Total	100	525	9	634

Table 3.2 - A summary table showing the age group and B-NHL subtypes of samples with copy number data.

3.3.3 Genomic abnormalities in B-NHL associated with age

Investigation of copy number aberrations in adult and paediatric B-NHL showed distinct differences between the two age groups. 410 samples with clinical data regarding age and disease type were analysed together using Nexus Copy Number 10.0 software and split into four groups based on diagnosis and age. A comparison of age and subtype diagnosis is displayed in Figure 3.1 and summarised in Table 3.2. Common abnormalities in paediatric Burkitt lymphoma included gain of 1q (24/49, 49.0%), amplification of 13q31.3 (14/49, 28.6%) and 7q gain (16/49, 32.7%).

The most frequently observed abnormalities in adult DLBCL were deletion of 6q (145/316, 45.9%), gain of chromosome 3 (103/316, 32.6%), gain of chromosome 18 (82/316, 25.9%) and *REL* amplification at 2p15 (52/316, 16.5%). Other frequently identified copy number aberrations were 9p21.3 loss, gain of chromosome 11q, gain of chromosome 12, deletion of 15p and deletion of 17p.

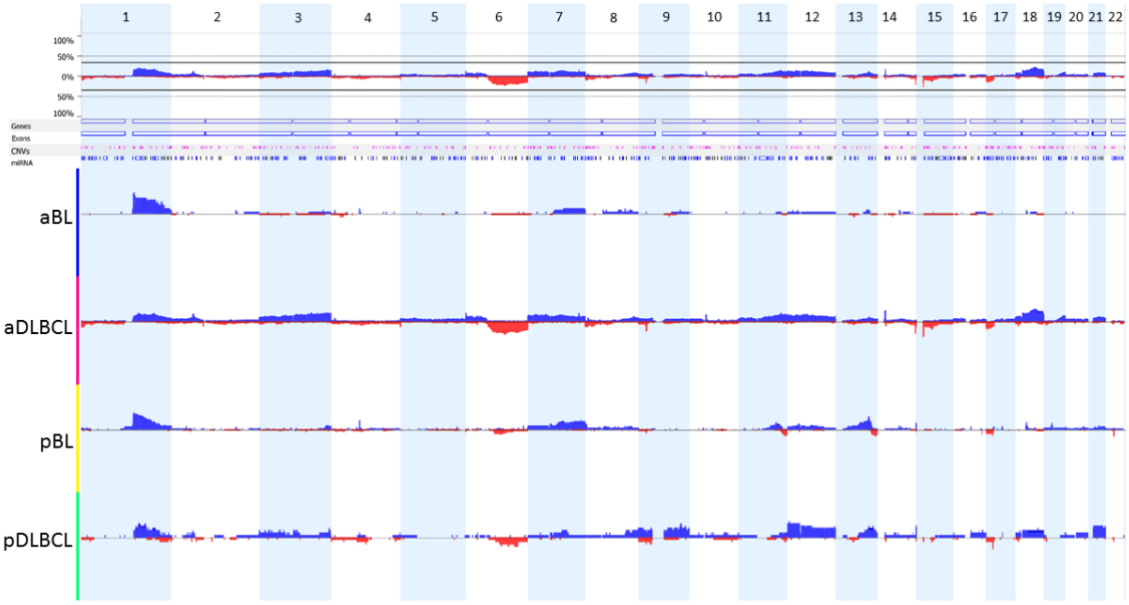


Figure 3.1 - Copy number profile for B-NHL subtypes separated by age and analysed by array. Blue peaks represent copy number gain and red peaks represent copy number loss at the region. A higher peak represents a higher incidence of abnormality. aBL = adult Burkitt Lymphoma, aDLBCL = adult Diffuse Large B Cell Lymphoma, pBL = paediatric Burkitt Lymphoma, pDLBCL = paediatric Diffuse Large B Cell Lymphoma.

The GISTIC1.0 algorithm implemented into Nexus Copy Number 10.0 identified 42 regions of significantly altered copy number in 634 B-NHL samples. 16 regions were

removed from analysis due to overlapping 100% with known CNVs and a further 7 were removed as they appeared to be caused by anomalous artefacts at the probe level. It was common for regions of copy number gain or loss to be identified by GISTIC1.0 that had low or uneven probe coverage, leading to inappropriate segmentation. The remaining 19 regions are presented in Table 3.3.

Cytoband	Event	G-Score	% of CNV Overlap	Genes of interest within region
1q23.3	CN Gain	52.2	0	*PBX1*
1p36.32	CN Loss	30.7	0	*TP73*
2p16.1	CN Gain	38.3	0	*REL*
4q21.22	CN Loss	22.2	10.4	None
6q21	CN Loss	66.3	0	*PDSS2, FOXO3, PRDM1*
7q21.3	CN Gain	40.6	18.0	*Mir106b-25*
7q12.2	CN Gain	32.2	0	*ABCA13, CDC14*
9p21.3	CN Loss	59.9	100	*CDKN2A, CDKN2B*
11q23.3	CN Gain	36.7	0	*CD3, KMT2A, TMPRSS4, IL10RA*
11p13	CN Gain	23.8	0	*CD44*
11q25	CN Loss	21.2	40.5	None
12q15	CN Gain	36.1	0	*IFN-γ, MDM1, IL26, IL22*
12p11.23	CN Gain	24.8	0	*ITPR2, SOX5*
13q31.3	CN Gain	34.6	6.2	*MIR17HG*
14q11.2	CN Gain	30.7	4.1	*SUPT16H, SNORD9, SNORD8, CHD8, RAB2B, TOX4, METTL3, SALL2*
15q15.1	CN Loss	27.1	3.1	*SNF106, CAPN3*
17p11.2	CN Loss	41.4	100	*TP53*
18q21.32	CN Gain	55.2	4.8	*CCBE1*
19p13.3	CN Loss	28.5	12.0	*CD70, VAV1, SH2D3A*

Table 3.3 – Regions and genes identified by GISTIC 1.0 analysis of all 634 B-NHL cases. Genes of interest were selected using Kegg Ontology Mapper (Kanehisa et al., 2019).

Additionally, GISTIC1.0 analysis was performed on a subset of the cohort comprised of 49 paediatric BL cases to identify recurrent abnormalities specific to pBL (Table 3.4).

Cytoband	Event	G-Score	% of CNV Overlap	Genes of interest within region
7q21.3-7q22.1	CN Gain	3.3	23.7	*MIR106B*
11q23.2	CN Gain	3.2	1.8	*HTR3A*, *HTR3B*
12q15	CN Gain	2.7	6.4	*MDM2*, *FRS2*
13q31.3	CN Gain	6.2	65.0	*MIR17HG*
17p11.2	CN Loss	5.8	100	* *TP53*
18q21.2	CN Gain	2.5	54.5	* *POLI*, *MBD2*, *C18ORF54*, *DYNAP*, *RAB27B*, *CCDC68*

*Table 3.4 - GISTIC 1.0 analysis of paediatric Burkitt Lymphoma samples. *Gene situated within wide peak region. Genes of interest were selected using Kegg Ontology Mapper (Kanehisa et al., 2019)*

In paediatric BL, 36 regions of copy number change were identified to be significant by the algorithm. Of these, 29 were excluded as they were due to array-specific

artefacts, regions of known germline copy number variation or were situated on sex chromosomes. The remaining 6 regions are summarised in table 5. Previously reported abnormalities including chromosome 1q gain, 7q21-q22 gain (*mir106b-25* locus), 13q31.3 amplification (*MIR17HG* locus) and *TP53* loss were identified by the algorithm. Additionally, copy number gain of *MDM2* and *FRS2* on chromosome 12q was reported in the minimal common region of 12q gains in paediatric BL.

Table 3.5 summarises the incidence of recurrent genomic abnormalities in B-NHL by subtype and age, demonstrating the difference in genomic landscape between paediatric and adult BL and DLBCL.

	aBL events	aBL %	pBL events	pBL %	Fisher exact *p* value	aDLBCL events	aDLBCL %	pDLBCL events	pDLBCL %	Fisher exact *p* value	Observation
1q gain	14/26	53.80%	24/49	49.00%	0.809	119/316	37.70%	9/19	47.40%	0.468	
REL amp	0/26	0%	0/49	0%	1	52/316	16.50%	2/19	10.50%	0.749	
Chr3 gain	4/26	15.40%	7/49	14.30%	1	103/316	32.60%	7/19	36.80%	0.802	
6q loss	1/26	3.90%	9/49	18.40%	0.150	145/316	45.90%	4/19	21.10%	0.055	
7q gain	3/26	11.50%	16/49	32.70%	0.054	92/316	29.10%	3/19	15.80%	0.296	
Chr9 gain	3/26	11.50%	6/49	12.20%	1	54/316	17.10%	7/19	36.80%	0.058	
CDKN2A loss	0/26	0%	4/49	8.20%	0.291	66/316	20.90%	3/19	15.80%	0.774	
11q gain	2/26	7.70%	12/49	24.50%	0.119	94/316	29.70%	3/19	15.80%	0.297	
Chr12 gain	3/26	11.50%	12/49	24.50%	0.234	98/316	31.00%	9/19	47.40%	0.203	
MIR17HG amp	2/26	7.70%	14/49	28.60%	0.042	29/316	9.20%	5/19	26.30%	0.033	Significantly more common in paediatric cases than adult.
13q34 loss	1/26	3.90%	7/49	14.30%	0.249	17/316	5.30%	0/19	0%	0.611	
17p loss	1/26	3.90%	7/49	14.30%	0.249	74/316	23.40%	5/19	26.30%	0.783	
Chr18 gain	2/26	7.70%	5/49	10.20%	1	72/316	22.80%	1/19	5.30%	0.087	

Table 3.5 - Incidence of frequent abnormalities in B-NHL stratified by patient age and disease subtype.

Amplification of the *REL* gene was not detected in BL and was found at a higher incidence in adult DLBCL than paediatric DLBCL. This observation is expected as REL amplification is typically seen in GCB-like DLBCL and not ABC-like DLBCL (Lenz *et al.*, 2008). Amplification of *REL* was identified in 2/19 pDLBCL cases in contrast to a previously published study (Klapper *et al.*, 2012) in which no paediatric patients aged below 15 years of age harboured this abnormality. The two pDLBCL cases with *REL* amplification were 7 and 14 years old. However, incidence of *REL* amplification in aDLBCL in our investigation was 16.14%, consistent with 17% observed in published data (Rosenwald *et al.*, 2002). Other abnormalities previously reported as being associated with DLBCL rather than BL were, as expected, observed at higher incidences in DLBCL including chromosome 3 gain, 6q loss, *CDKN2A* loss, and chromosome 18 gain. 6q loss and chromosome 18 gains were shown to be more frequent in adult than paediatric DLBCL. The genomic landscape of adult BL appeared similar to that of the paediatric subtype, but with fewer additional abnormalities (Figure 3.1) and lower incidences of reported recurrent abnormalities (Table 3.5) with the exception of 1q gain. Chromosome 12 gains were reported at a higher frequency in DLBCL than BL but also at a higher incidence in paediatric than adult disease in both subtypes.

MIR17HG amplification was identified as a significant abnormality by our GISTIC analysis in BL and it has been shown recently to be more common in paediatric disease than adult disease (Havelange *et al.*, 2016b). Our analysis determined that the incidence of *MIR17HG* amplification is higher in paediatric DLBCL as well as in BL. This suggests that it is a paediatric-associated abnormality in B-NHL regardless of disease subtype. 14/49 (28.6%) paediatric BL patients and 5/19 (26.3%) paediatric DLBCL cases harboured a gain of *MIR17HG*. In contrast, the incidence of *MIR17HG* amplification was significantly lower in adult patients, at 2/26 (7.7%) and 29/316 (9.2%) in adult BL and DLBCL respectively (BL p=0.042, DLBCL p=0.033, Fisher exact test) (Figure 3.2). This suggests that *MIR17HG* may be useful for creating a biologically relevant cut off for paediatric and adult type disease.

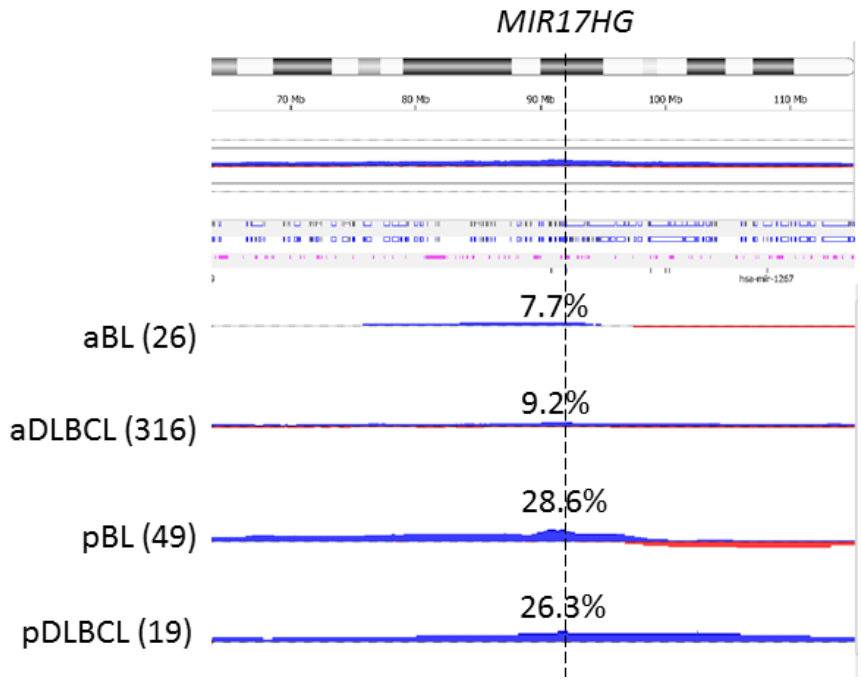


Figure 3.2 - Incidence of MIR17HG amplification in adult and paediatric BL and DLBCL. Blue denotes copy number gain and red denotes copy number loss. The higher the amplitude of the peak, the higher the frequency of aberration.

3.3.5 Teenagers and Young Adults

The age range of 15 and 24 years of age was used to define the teenager and young adult subset of patients. This was based on results from other studies that defined 15-18 as adolescent patients and 24 as the age at which abnormalities common in adult patients become more frequent (Klapper *et al.*, 2012). This group of patients had generally quiet genomes with recurrent gains of 1q and chromosome 12. Of the 17 patients aged 15-24, four out of 17 patients had gains of 12p and 12q, with an additional patient harbouring gain of 12q alone. Two patients, aged 17 (MPI-757) and 18 (MPI-122), harboured gains of chromosomes 9 and 12 together. MPI-757 also had a 1.5Mb deletion on 3p that involved *RHOA* and a 1.8Mb amplification involving *MIR17HG*. MPI-122 had only chromosome 9 and 12 gains and a small deletion on 2p involving *DCDC2C and ALLC*. The rest of the cohort with age data was then screened for this combination of whole chromosome gains. Twelve cases aged between 7 and 87 were identified with copy number gains of at least one chromosome arm of both chromosome 9 and 12.

By splitting all B-NHL samples with patient age available into age groups spanning 5 years it was possible to visualise the change in the genomic landscape. Using this method, a subset of patients between the ages of 16 and 20 years of age were

identified with gains of chromosome 9 and chromosome 12 concurrently. This combination of abnormalities is previously unreported in B-NHL, and is present at a much higher incidence in this age group (Figure 3.3).

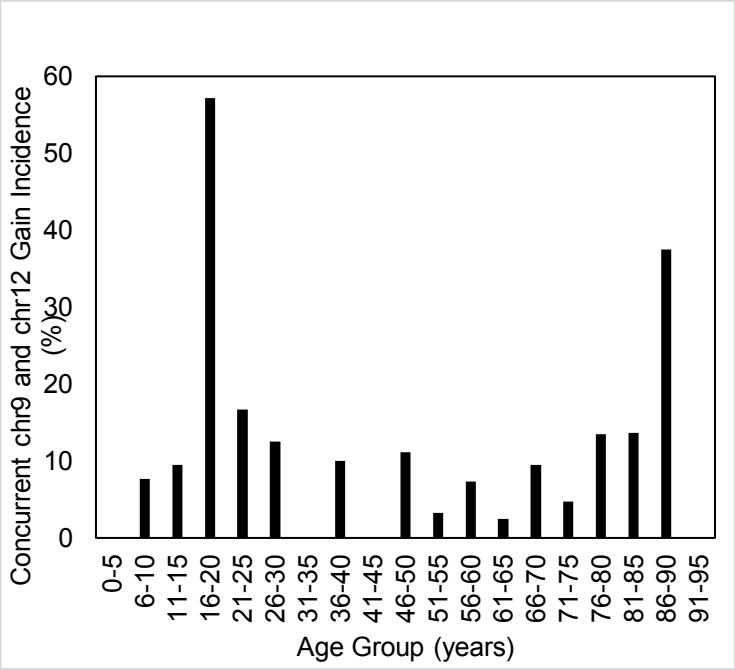


Figure 3.3 - Incidence of concurrent chromosome 9 and chromosome 12 gains in all B-NHL cases stratified by age.

This analysis also confirmed that the frequency of *MIR17HG* amplification was lower in adult patients (Figure 3.4).

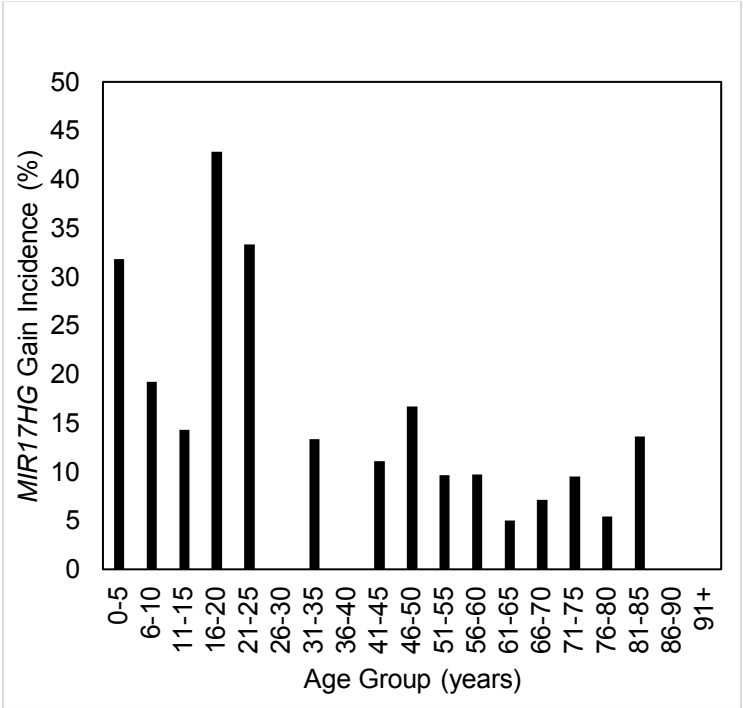


Figure 3.4 - Incidence of *MIR17HG* amplification in all B-NHL cases stratified by age.

As this analysis shows that the incidence of *MIR17HG* gain is more frequent in children than adults, *MIR17HG* may be used as a marker to define a cut-off age between paediatric and adult B-NHL. A Chi-squared test showed that the age cut-off with the most significant difference in incidence of *MIR17HG* amplification was age 25.

3.3.6 Genomic abnormalities associated with outcome

In total, 292 of 634 B-NHL patients analysed in this study had associated outcome and follow up data allowing for analysis of genomic aberrations by patient survival. Of 49 pBL cases identified in this cohort there were 36 patients with follow up time and outcome information. By integrating this clinical data with copy number data it was possible to evaluate the impact of abnormalities on patient survival. Of the six frequently observed abnormalities identified by GISTIC analysis (Table 3.4), 17p loss was associated with a significantly altered outcome by Cox proportional hazard method (HR 7.7 95% CI 1.3-46.0, p=0.026, Cox proportional hazard method). 17p loss was associated with an inferior overall survival at 3 years, demonstrated in Figure 3.5. There were insufficient cases for analysis of time to progression or association with relapse/refractory disease.

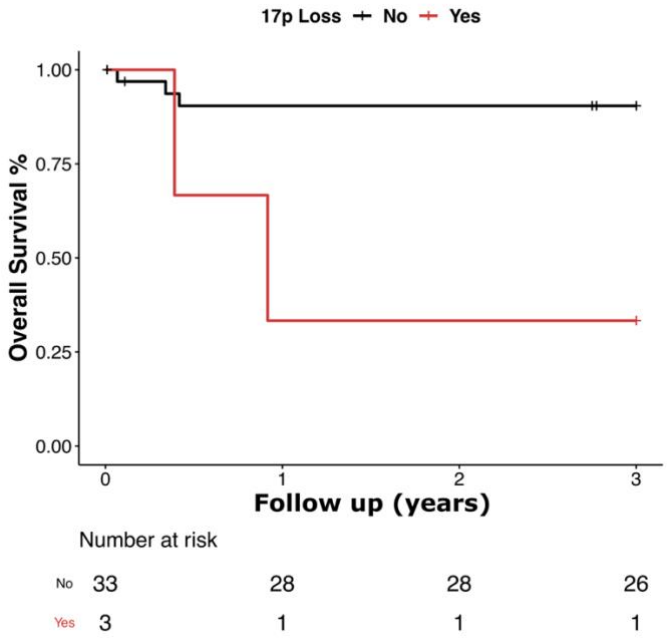


Figure 3.5 - Kaplan-Meier analysis of the prognostic significance of 17p deletion in 36 paediatric Burkitt Lymphoma patients. (p=0.009, Log rank test)

The most well-characterised tumour suppressor gene, *TP53*, is situated at 17p13, which was deleted in all cases with 17p loss. These preliminary findings suggest that the *TP53* gene is important in disease progression in paediatric BL.

Kaplan Meier analysis was performed on the subset of BL patients with outcome data to determine whether previously identified prognostic markers were indicative of poor outcome in this cohort. In keeping with the analysis performed by Schiffman *et al*, gain of 13q31.3, involving *MIR17HG* was associated with a poorer outcome (p=0.024, Log rank test) which seemed to validate the results of the previous Schiffman study (Schiffman *et al.*, 2011) (Figure 3.6).

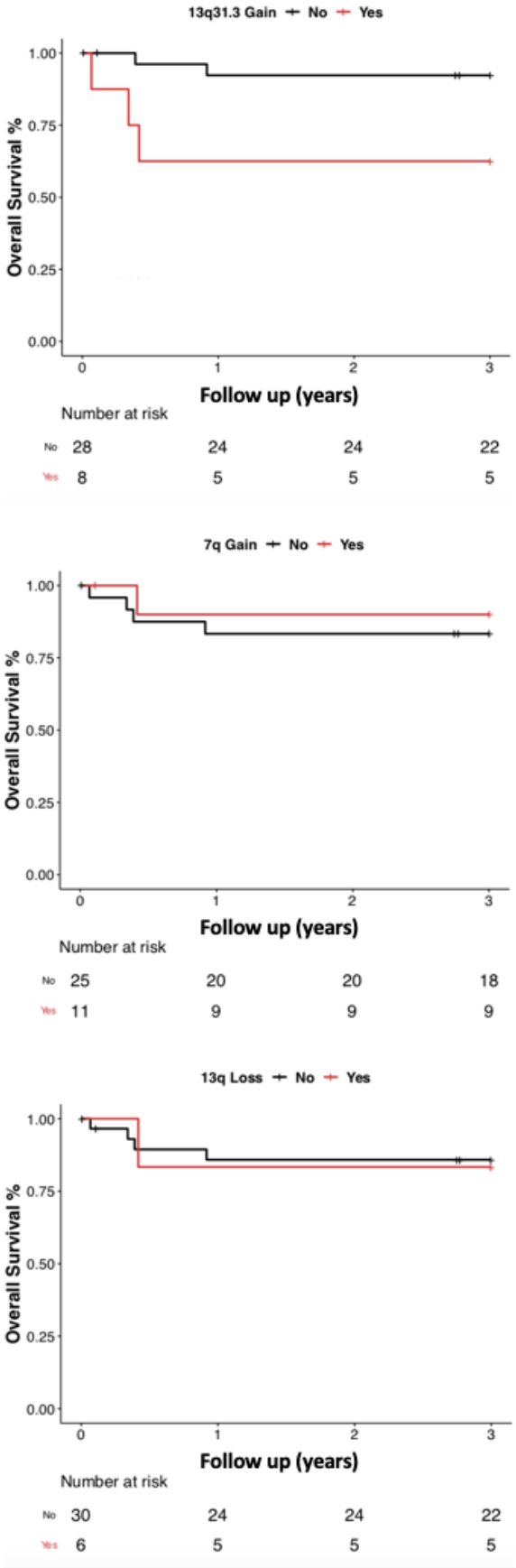


Figure 3.6 - Kaplan Meier Analysis was performed to evaluate the effect of previously identified putative prognostic markers on overall survival in paediatric BL. (P values: p=0.024, p=0.59 and p=0.91 respectively, Log rank test).

7q gain was not associated with an inferior overall survival (p=0.59, Log rank test). Equally, deletion of 13q, as reported to be prognostic in the Poirel (2009) and Nelson (2010) studies was not associated with poorer outcome (p=0.91, Log rank test). This apparent disparity in prognostic power between our analysis and those previously published may be explained by the small sample size in this analysis. Investigation with an expanded cohort will allow for more comprehensive analysis of putative prognostic markers.

3.4 Discussion

Initial copy number investigation utilising existing data from in-house datasets successfully identified recurrent copy number abnormalities that were enriched in age-associated subgroups. Analysis of copy number array data from 634 B-NHL patients has enabled detailed analysis of genomic abnormalities in the disease. Abnormalities reported in the literature in Burkitt Lymphoma including gain of 1q, gain of 7q and 13q abnormalities were identified in this investigation. We report gain of 1q in 49% of paediatric BL cases, a higher incidence than observed in the Poirel study (Poirel *et al.*, 2009). This difference in frequency is most likely explained by increased sensitivity of array platforms compared to traditional karyotyping and the ability to detect smaller regions of copy number change. Gain of 7q was detected in 32.7% of paediatric BL which is also considerably higher than the 10% reported by Nelson *et al*. (2010) and Poirel *et al*. (2009). Deletion of 13q, in contrast, was found in a lower proportion of paediatric BL cases than has been previously reported by Nelson *et al*.(2010). We detected 7 cases (14.3%) with deletions on 13q and all involved 13q34. No deletions of 13q14.3 were detected. The frequency of 13q deletions in our study correlates with the findings of Poirel *et al*. (2009). No adult-specific abnormalities were detected in BL and samples in this group appeared to have much quieter genomes than the paediatric cases. Loss of 6q and gain of chromosome 18 were both most strongly associated with adult DLBCL and were observed at lower incidences in BL. This reflects the increase in ABC-subtype DLBCL with age reported in the literature (Deffenbacher *et al.*, 2012).

Our copy number analysis confirmed the finding reported by Havelange *et al*.(2016b) that *MIR17HG* amplification is found at higher frequencies in paediatric Burkitt lymphoma than in the adult disease. Remarkably, investigating paediatric cases and separating out disease subtypes revealed that *MIR17HG* is associated with patient age in B-NHL regardless of subtype, with similar incidences in paediatric BL and DLBCL, and adult BL and DLBCL respectively. Stratifying patients into 5-year age groups showed the trend of *MIR17HG* amplification being at a high frequency in paediatric cases before reducing in frequency after age 25. We did not, however, find *MIR17HG* in paediatric and young adult cases exclusively, as we detected *MIR17HG* gain in 29/316 (9.2%) adult DLBCLs up to the age of 84 years. The oldest BL patient with *MIR17HG* gain was 35, suggesting that the abnormality may not occur in older adults. This would need validation with a larger cohort of BL patients however.

Despite data from previous studies (Klapper *et al.*, 2012) amplification of *REL* was detected in both adult and paediatric DLBCL cases at similar frequencies, suggesting that the abnormality is not age specific. Gain of chromosome 9 and gain of chromosome 12 were both observed as being most frequent in paediatric DLBCL. Three paediatric DLBCL patients harboured both abnormalities concurrently, and two these patients were adolescents aged 17 and 18. Additional data from paediatric and TYA cases is needed to further investigate whether these co-occurring events are significant.

Identification of a potential TYA-specific abnormality suggests that a subgroup of patients in this age group may be biologically distinguished from paediatric and adult type B-NHL. Further investigation with a higher number of patients between 16 and 30 may allow us to determine which genes and pathways are involved with a view to developing a novel therapeutic strategy for this group. Unfortunately, as diagnosis of patients at this age group is very low, it is difficult to find enough samples for meaningful analysis with statistically significant results. Klapper *et al.* (2012) suffered similar setbacks, as in their cohort of 364 patients, only 5 fell within their “adolescent” cut-off. Our investigation had 7 patients aged between 16 and 20 years of age with strikingly quiet genomes when compared to typical adult and paediatric B-NHL cases with the notable exception of concurrent chromosome 9 and 12 gains. TYA cases are under-represented in this and other studies due to the reduced incidence of disease in these patients, primarily as a result of fewer cases being diagnosed in these age groups. Meaningful characterisation of the TYA genomic landscape in B-NHL would require more patients in this age group with array data and/or next-generation sequencing data.

Deletion of 17p involving *TP53* was shown to be significantly associated with a poorer overall survival in paediatric BL. However, no other abnormalities were significantly associated with a poor prognosis. Although reported by Poirel *et al.* (2009) there was no significant effect of 7q gain or 13q loss on outcome in paediatric *BL* in our combined analysis. While *MIR17HG* gain was shown to be a prognostic marker in paediatric BL as part of our study, this was only in a very small subgroup of BL cases with outcome. This analysis highlighted the need for a larger cohort for meaningful survival analysis with better annotated clinical data. The majority of B-NHL data available in the literature was either from adult studies or were of an

unknown age, and no further copy number data from paediatric B-NHL was identified.

The primary aim of the study summarised in this chapter was to determine whether combining existing in-house data could be analysed in combination with previously published data generated on numerous different array platforms. There were more samples available from the CCLG cohort for paediatric B-NHL, so in light of the low numbers in this study 110 cases were sent to AROS (Applied Biosystems, Denmark) for analysis with Affymetrix Cytoscan HD array kit and Affymetrix Oncoscan FFPE array kit. However, many cases had not had their pathology reviewed to confirm the diagnosis from the treatment hospital. This is a significant issue in paediatric B-NHL as the subtype is not used to drive treatment decisions currently; DLBCL and BL patients all receive FAB/LMB96 protocol chemotherapy, as do newer subtypes including HGBL, NOS and BLL-11q. The identification of prognostic genomic markers in paediatric B-NHL will likely reflect the underlying biology of different disease subtypes, making robust pathological diagnosis of subtype important. Indeed, newly defined subtypes of B-NHL that were included in the 2016 revision of the WHO Classification of Lymphoid Neoplasms (Swerdlow *et al.*, 2016) were not in use when many of these cases were diagnosed. Great effort was made to perform a thorough pathology review led by pathologist Dr Chris Bacon, which successfully confirmed the diagnosis of 74/110 (67.2%) patients in the new cohort. Lack of FFPE material was the primary reason for path review not being completed. On top of this, additional data requests were made to the CCLG for clinical data required for reliable survival analysis. Patient age, gender, bone marrow involvement status, CNS involvement status, CSF involvement status, date of diagnosis, date of end of treatment, date of relapse, relapse status, date of last follow up, treatment type, FAB/LMB96 treatment group and rituximab treatment status were requested from CCLG centres for cases in the study. Response rates from centres were high, and we received sufficient data for 101/110 CCLG patients to perform survival analysis for overall survival and time to progression. This cohort of 110 patients was combined with the 52 paediatric cases identified from the literature during this pilot project. This new combined paediatric cohort was 162 patients was taken forward for a more comprehensive analysis detailed in Chapter 4.

Chapter 4: Copy Number Analysis of Paediatric B cell non-Hodgkin Lymphoma

4.1 Introduction

IG-MYC rearrangement is the hallmark of BL and is observed in 90-95% of BL cases (van Rijk *et al.*, 2008). However, the co-operating cytogenetic and molecular genetic abnormalities are less well defined in BL, particularly those associated with outcome. Early cytogenetic studies using traditional karyotyping reported that additional copies of chromosome 1q and 7q were common in BL patients, as well as deletion of 13qter and 17p (Lones *et al.*, 2004). However, most of these studies were conducted on small numbers of patients and/or used early, lower resolution CGH arrays. Cohort size is a major hurdle to overcome in the study of prognostic markers in B-NHL, which is a rare disease in the paediatric setting. Small patient numbers limit our ability to understand the impact of B-NHL genomic abnormalities and limit the significance of statistical analysis.

With the advent of high resolution copy number arrays several studies were published investigating copy number changes in BL (Scholtysik *et al.*, 2010) and DLBCL (Scholtysik *et al.*, 2015). In BL the most common aberrations reported were gain of 1q, gain of 7q and gain of 13q31.3 with deletions on 6q and 17p. In DLBCL, 1q gain, chromosome 7 gain and 6q loss were also recurrent while abnormalities on chromosome 3 and 18 were more common than in BL. ABC- and GCB-like DLBCL had distinct copy number patterns, with gain of chromosomes 3 and 18 and loss of 6q involving *BACH2* being more common in the ABC-DLBCL subtype while *REL* amplification has been frequently shown to occur in GCB-like DLBCL (Scholtysik *et al.*, 2015). *CDKN2A* deletion was significantly more frequent in ABC-like DLBCL. As *CDKN2A* deletion has been shown to be associated with poorer outcome in adult DLBCL, this may go some way to explain the inferior survival rates of ABC-like DLBCL compared to the GCB-like subtype (Jardin *et al.*, 2010). In contrast, *CDKN2A* deletion is seen in BL in a small population of patients, but has not been shown to be associated with poor outcome (Scholtysik *et al.*, 2010; Schiffman *et al.*, 2011). However, these studies involve primarily adult patients, which highlights the importance of clinical context to prognostic markers in B-NHL. Treatment protocols differ between BL and DLBCL subtypes in the adult setting, but not in paediatrics, where patients are primarily treated with FAB/LMB96 protocols (Patte *et al.*, 2007). It

remains to be seen whether prognostic markers identified in BL would apply to DLBCL, and vice versa.

Chemotherapy resistance represents the most pressing challenge in treating B-NHL and clinically actionable prognostic markers are urgently needed. A retrospective study of karyotypes from 33 children and 37 adults with BL suggested that abnormalities involving 13q and 17p were associated with poor outcome in children and adults respectively (Onciu *et al.*, 2006). Low resolution cytogenetic analysis of 238 patients enrolled on the FAB/LMB trial revealed that additional copies of chromosome arm 7q and del(13q) were associated with a significantly poorer EFS in BL and DLBCL patients (Poirel *et al.*, 2009). Subsequent work using copy number arrays have reported amplification of the *MIR17HG* miRNA cluster on chromosome arm 13q as a putative prognostic marker, observed in 3/28 patients (10.7%) (Schiffman *et al.*, 2011). However, this study was conducted on only 28 BL patients with only three relapse patients and was not powered to identify markers of relapse. Despite putative prognostic markers being identified as early as 2009, none of these have been translated to the clinic or been the focus of clinical trials. A significant contributor to this disconnect between research and the clinic is the relative rarity of paediatric B-NHL. Traditional clinical trial designs are unrealistic in paediatric B-NHL as recruitment would need to run for years. Similarly, small patient cohorts in genomic studies are not sufficient to change opinions in the field regarding treatment, meaning that larger studies are required. This was achieved in this study by collecting data from the literature to complement the cases from our in- house cohort. This facilitated creation of the largest paediatric B-NHL cohort with copy number and clinical data which was analysed to determine whether genomic aberrations detectable by copy number array can be used to stratify patients at the point of diagnosis and used to direct clinical management of the disease.

4.2 Chapter Aims

The specific aims of the work detailed in this chapter were as follows:

- Identify and characterise copy number aberrations in a large paediatric B-NHL cohort using integrated bioinformatics.
- Evaluate the prognostic value of recurrent genomic abnormalities by univariate and multivariate modelling.

- Further investigate statistically significant prognostic genomic markers using whole exome sequencing data.

4.3 Results

4.3.1 Clinical Characteristics of the Cohort

Copy number analysis of 162 paediatric B-NHL samples was performed to identify recurrent genomic abnormalities. The cohort comprised 109 BL, 36 DLBCL, 5 BLL-11q, 1HGBL, NOS and 11 B-NHL, NOS. Of these, 110 cases were acquired from the CCLG Tissue Biobank and the remaining 52 were taken from eight existing studies in the literature as described in Chapter 3 (Lenz *et al.*, 2008; Kato *et al.*, 2009; Lepretre *et al.*, 2010; Scholtysik *et al.*, 2010; Kwiecinska *et al.*, 2014; Suguro *et al.*, 2014; Scholtysik *et al.*, 2015; Havelange *et al.*, 2016b). The distribution of age, sex, disease stage and BM/CNS involvement were similar between cases with and without disease progression (Table 4.1).

		Outcome		
Characteristic		**No Progression**	**Disease Progression**	**Unknown**
Total Number of Cases		119	18	25
Sex				
	F	22 (18.5)	6 (33.3)	1 (4)
	M	93 (78.2)	11 (61.1)	7 (28)
	U	4 (3.4)	1 (5.6)	17 (68)
Age at Diagnosis				
	0-5	29 (24.4)	7 (38.9)	6 (24)
	6-10	40 (33.6)	7 (38.9)	12 (48)
	11-14	35 (29.4)	3 (16.7)	3 (12)
	15-18	15 (12.6)	1 (5.6)	4 (16)
	Median	8	6.5	8
	Range	2-18	0-16	2-18
Tumour Stage				
	1	2 (1.7)	0 (0)	0 (0)
	2	12 (10.1)	1 (5.6)	0 (0)
	3	18 (15.1)	4 (22.2)	0 (0)
	4	10 (8.4)	1 (5.6)	1 (4)
	U	77 (64.7)	12 (66.7)	24 (96)
BNHL Subtype				
	BL	84 (70.6)	15 (83.3)	10 (40)
	DLBCL	22 (18.5)	2 (11.1)	12 (48)
	BLL-11q	4 (3.4)	0 (0)	1 (4)
	HGBL, NOS	1 (0.8)	0 (0)	0 (0)
	B-NHL, NOS	8 (6.7)	1 (5.6)	2 (8)
CSF Involvement				
	Y	2 (1.7)	0 (0)	0 (0)
	N	60 (50.4)	10 (55.6)	2 (8)
	U	57 (47.9)	8 (44.4)	23 (92)
CNS Involvement				
	Y	7 (5.9)	1 (5.6)	0 (0)
	N	73 (61.3)	11 (61.1)	2 (8)
	U	39 (32.8)	6 (33.3)	23 (92)
BM Involvement				
	Y	11 (9.2)	4 (22.2)	1 (4)
	N	58 (48.7)	9 (50.0)	1 (4)
	U	50 (42.0)	5 (27.8)	23 (92)
MYC translocation				
	MYC-R (IG)	75 (63.0)	15 (83.3)	8 (32)
	MYC-R (non-IG)	1 (0.8)	0 (0)	0 (0)
	No MYC-R	30 (25.2)	2 (11.1)	2 (8)
	NA	13 (10.9)	1 (5.6)	15 (60)
Treatment Protocol				
	FAB/LMB-based	86 (72.3)	14 (77.8)	2 (8)
	UKCCSG NHL-based	3 (2.5)	0 (0)	1 (4)
	R-CHOP	0 (0)	1 (5.6)	0 (0)
	ALL-like	14 (11.8)	1 (5.6)	0 (0)
	CHOP-like	1 (0.8)	0 (0)	2 (8)
	NHL BFM-90	1 (0.8)	0 (0)	1 (4)
	Unknown	13 (10.9)	2 (11.1)	19 (76)
	None	1 (0.8)	0 (0)	0 (0)
Rituximab Treatment				
	Y	4 (3.4)	6 (33.3)	1 (4)
	N	58 (48.7)	4 (22.2)	1 (4)
	U	57 (47.9)	8 (44.4)	23 (92)
Dead*				
	N	104 (87.4)	5 (27.8)	13 (52)
	Y	6 (5.0)	13 (72.2)	4 (16)
	U	9 (7.6)	0 (0)	8 (32)
	Median time to death (mo)	1.4	7	4.38
	Range (mo)	0.07-5.00	3.58-17.05	1.87-24.0

Table 4.1 - Clinical characteristics of 162 paediatric B-NHL patients. F=Female, M=Male, U=Unknown, BL=Burkitt lymphoma, DLBCL=Diffuse large B-cell lymphoma,

*BLL-11q=Burkitt-like Lymphoma with 11q aberration,HGBL,NOS=High Brade B-cell Lymphoma not otherwise specified, B-NHL,NOS= B-cell non-Hodgkin lymphoma not otherwise specified, N= No, Y=Yes, MYC-R=MYC rearrangement, IG=Immunoglobuli, NA=not available. * censored at 3 years follow-up.*

18 cases had disease progression and 119 were in remission without progressive disease and in total, 128/162 (79%) had sufficient follow up data for survival analysis with a median follow up time of 5 years (range 0.01-22.5 years). Of the 18 cases with progressive disease, 12 patients achieved remission and six did not achieve remission prior to disease progression. There was no statistical significance between disease progression and subtypes (BL 15/109 13.8% vs 2/36 DLBCL 5.6%; p=0.24, Fisher exact test). Notably, all disease progression occurred within one year of original diagnosis (mean 5.01 months, range 2.80-8.25 months).

4.3.2 Analysis of Copy Number Abnormalities

The genomic profiles of 109 BL, 36 DLBCL, 5 BLL-11q, 1HGBL, NOS and 11 B-NHL, NOS cases were analysed (Figure 4.1). The mean number of copy number abnormalities (CNAs) and regions of copy-number neutral loss of heterozygosity (CNN-LOH) per sample were 40.43 (range: 0-618) and 1.72 (range: 0-30), respectively. Samples from the CCLG cohort were processed and segmented using Affymetrix arrays as detailed in Chapter 2. Of the 110 CCLG cases in this study, 79 were analysed using the Affymetrix Cytoscan HD array kit, 26 on the Affymetrix Oncoscan 2.0 MIP array kit and 5 on the Affymetrix Genome-Wide Human SNP6.0 array kit. Segmentation was performed using the proprietary Nexus Copy Number 10 software. Affymetrix Cytoscan HD and Genome-Wide Human SNP6.0 array cases were segmented using the SNP Rank algorithm incorporated in Nexus Copy Number 10, which is based on the Circular Binary Segmentation (CBS) algorithm. Affymetrix Oncoscan MIP arrays were segmented using the TuScan algorithm. All array types were segmented to generate calls for copy number gain, deletion and CNN-LOH.

Copy number calls were manually checked for over-segmentation, noise and recurrent germline CNVs. Over-segmentation occurs when the threshold set for the segmentation algorithm is close to the probe median of a region of copy number change, resulting in uneven segmentation across the true abnormality (Rieber *et al.*, 2017). Suspected germline CNVs were checked against the Database of Genomic Variants (MacDonald *et al.*, 2014).

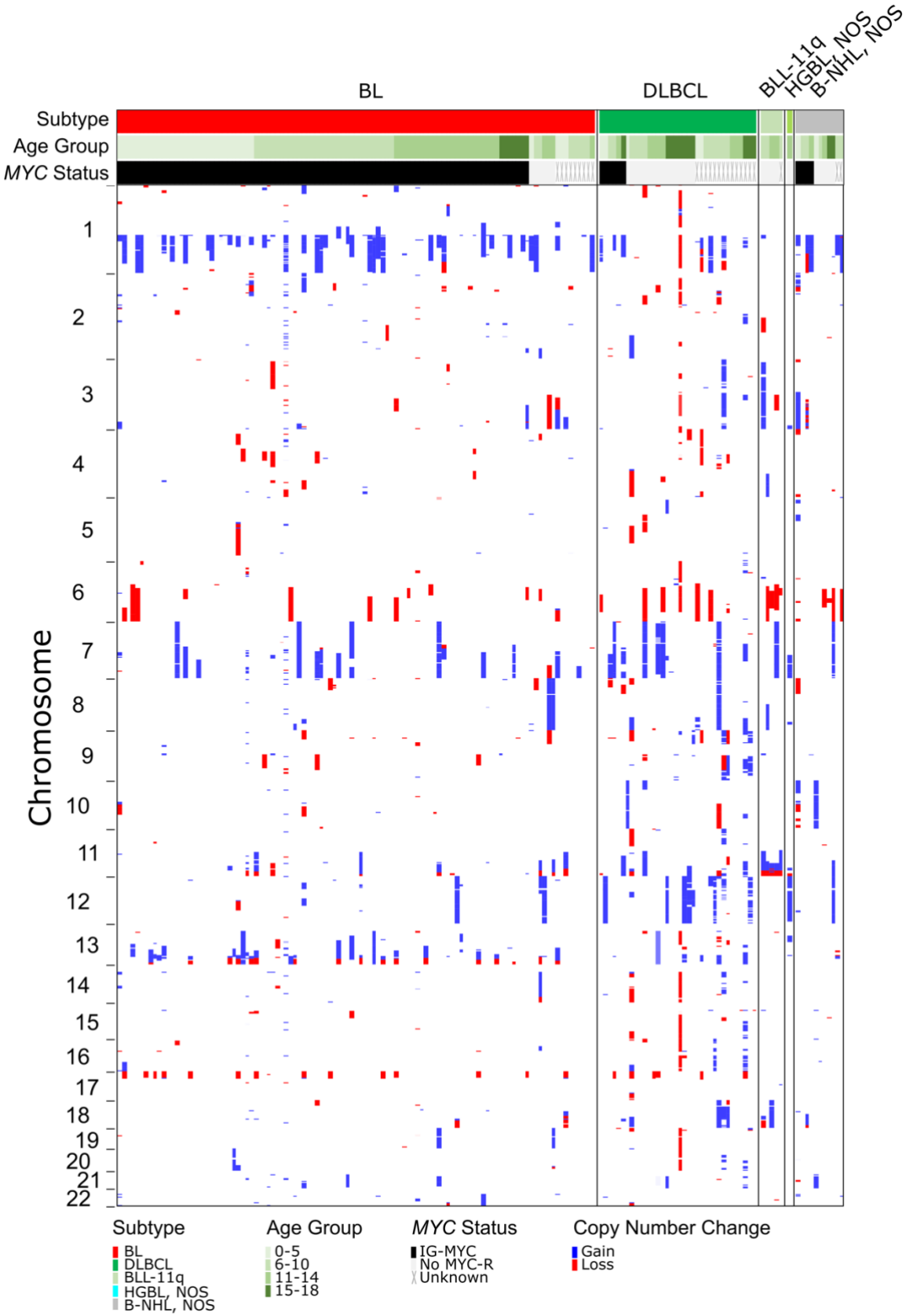


Figure 4.1 - Heatmap displaying segmented copy number data for 162 paediatric B-NHL patients. Age group, MYC status and diagnosis are shown. Sex chromosomes were excluded from analysis.

4.3.3 GISTIC2.0 Analysis of Significant Targets

Significant regions of copy number alteration were detected using the GISTIC2.0 algorithm. GISTIC2.0 is the second iteration of a tool developed by the Broad Institute. The tool takes segmented copy number data and identifies regions that are likely to be driver events in cancer based on the frequency and amplitude of the aberration. This tool was utilised to highlight potential driver events in paediatric B-NHL which could then be evaluated through survival analysis to uncover potential prognostic markers and drug targets. In total, 26 focal recurrent regions of focal CNA or CNN-LOH were detected by the GISTIC2.0 algorithm in the 162 paediatric B-NHL cases identified; including eight regions of copy number gain, 10 of loss and eight of CNN-LOH (Figure 4.2A, Table 4.2).

cytobands	Abnormality	Size (mb)	q value	residual q value	wide peak boundaries
13q31.3	Gain	0.28	2.3163e-19	2.3163e-19	chr13:91778876-92062330
11q23.3	Gain	1.17	1.8438e-05	1.8438e-05	chr11:117745876-118920452
3q29	Gain	2.30	0.00015311	0.00015311	chr3:195718766-198022430
1p12-1q21.3	Gain	33.88	0.00078668	0.00078668	chr1:119258849-153139036
18q21.2-18q21.31	Gain	3.53	0.0077145	0.0077145	chr18:51649814-55175432
14q24.3	Gain	0.16	0.016323	0.016323	chr14:73869638-74029913
2p16.1-2p15	Gain	3.21	0.020563	0.020563	chr2:60281930-63491437
10q21.1-10q21.3	Gain	5.49	0.058425	0.058425	chr10:60382406-65875657
9p21.3	Loss	0.59	1.0925e-09	1.0925e-09	chr9:21860429-22448736
11q24.1-11q25	Loss	11.20	1.0925e-09	1.0925e-09	chr11:123801732-135006516
13q33.1-13q34	Loss	12.39	1.0925e-09	1.0925e-09	chr13:102776066-115169878
3q13.31-3q13.32	Loss	2.50	7.7413e-07	7.7413e-07	chr3:116150887-118646958
2p22.2-2p16.3	Loss	12.70	0.00018371	0.00018371	chr2:37887965-50591552
6q13-6q22.1	Loss	42.69	0.00082233	0.00082233	chr6:74521195-117212533
19p13.3-19p13.2	Loss	0.63	0.0034498	0.0034498	chr19:6470214-7103541
4q12-4q22.3	Loss	43.82	0.027997	0.028188	chr4:54669141-98489305
8p23.3-8p23.2	Loss	2.82	0.028448	0.028188	chr8:1-2822739
1q25.1-1q44	Loss	75.32	0.084717	0.077888	chr1:173932804-249250621
1p36.33-1p34.3	CNN-LOH	35.90	4.0217e-07	4.0217e-07	chr1:704575-36603389
17q22-17q24.1	CNN-LOH	10.03	0.00069264	0.0053824	chr17:53498472-63525564
1q25.3-1q44	CNN-LOH	68.79	0.0067764	0.0067595	chr1:180465542-249250621
9p24.3-9p21.3	CNN-LOH	21.80	0.034211	0.034211	chr9:199493-22001547
17q11.2-17q21.32	CNN-LOH	18.72	0.052759	0.20588	chr17:26213681-44931930
6p25.3-6q11.1	CNN-LOH	62.40	0.23951	0.23951	chr6:1-62399484
15q13.1-15q23	CNN-LOH	42.62	0.23951	0.23951	chr15:28331173-70954036
17p13.3-17p11.2	CNN-LOH	16.46	0.23951	0.23951	chr17:1-16462456

Table 4.2 - GISTIC2.0 Focal regions identified in 162 paediatric B-NHL patients

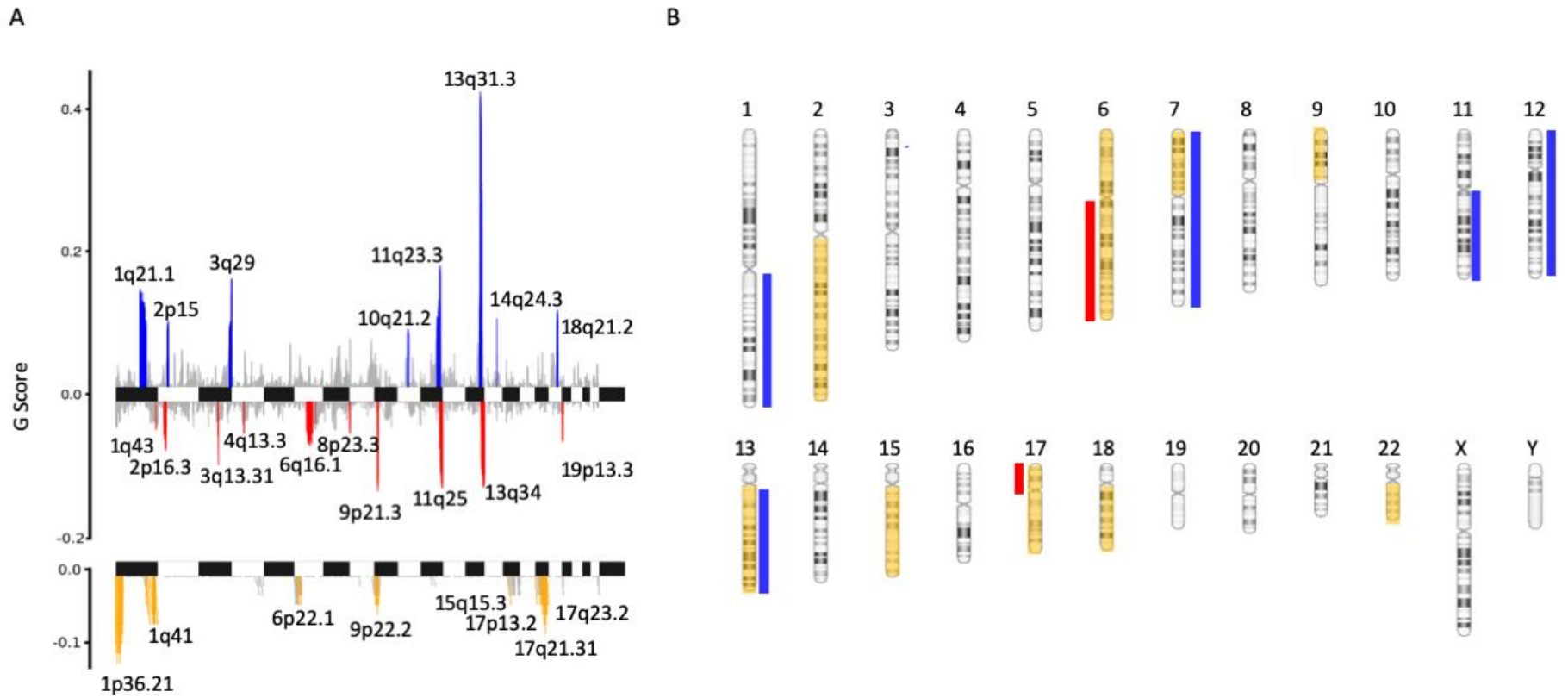


Figure 4.2 - GISTIC2 results for 162 paediatric B-NHL patients with copy number data. (A) Focal GISTIC2 scores representing recurrent smaller copy number aberrations. (B) - An ideogram showing recurrent broad arm-length aberrations. Blue denotes copy number gain, red denotes copy number loss and yellow denotes CNN-LOH.

GISTIC2.0 improves upon the previous iteration of the tool in its ability to separate out focal events and larger, arm length events. Figure 4.2A shows the focal events identified in the cohort of 162 paediatric B-NHL patients. Focal gain of 13q31.3 (0.28 Mb) and deletion of 9p21.3 (0.59 Mb), were the most GISTIC2.0 significantly gained and lost regions occurring in 25% (40/162) and 13% cases (21/162), respectively. Broad analysis of chromosome arm-length aberrations identified seven gained arms, two lost arms and 11 arms with recurrent CNN-LOH (Figure 4.2B, Table 4.3).

Arm	Abnormality	# Genes	Amp q-value
1q	Gain	1955	0
7q	Gain	1277	0
12p	Gain	575	7.82E-13
7p	Gain	641	5.28E-10
12q	Gain	1447	1.03E-08
13q	Gain	654	1.25E-07
11q	Gain	1515	0.00564
17p	Loss	683	0
6q	Loss	839	2.83E-07
6p	CNN-LOH	1173	0
9p	CNN-LOH	422	0
17p	CNN-LOH	683	0
17q	CNN-LOH	1592	0
13q	CNN-LOH	654	2.09E-08
22q	CNN-LOH	921	2.54E-06
18q	CNN-LOH	446	0.00488
15q	CNN-LOH	1355	0.0144
7p	CNN-LOH	641	0.0199
2q	CNN-LOH	1556	0.0251
6q	CNN-LOH	839	0.0475

Table 4.3 - GISTIC2.0 Broad regions identified in 162 paediatric B-NHL patients

4.3.4 GISTIC2.0 Focal Analysis to identify focal driver copy number abnormalities

The most significant region of gain detected by GISTIC2.0 was on 13q31.3 spanning 0.28Mb and involved the *MIR17HG* locus (Figure 4.3). This microRNA cluster gene has previously been studied in B-NHL and has been shown to be significantly more common in paediatric cases compared to adults (Havelange *et al.*, 2016b). Another

study identified amplification of *MIR17HG* in 2/3 relapse cases and suggested that the aberration was associated with disease progression (Schiffman *et al.*, 2011). The region also overlapped with the *GPC5* gene which codes for a cell surface proteoglycan, but evidence from the literature strongly suggests that *MIR17HG* is the region of interest within 13q31.3 copy number gain. Expression of *MIR17HG* polycistron was shown to synergise with *MYC* expression in nude mice to accelerate tumour growth (Tagawa *et al.*, 2007). Tumours that arose in mouse models of co-expression of *MYC* and *MIR17HG* were B cell lymphomas (He *et al.*, 2005).

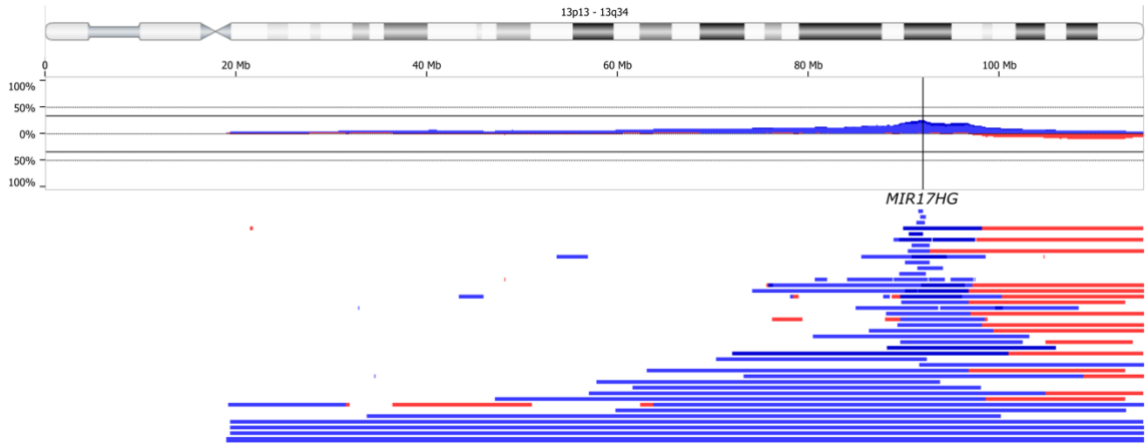


Figure 4.3. - A plot showing copy number data for 13q from 41 patients with 13q31.3 copy number gain. Blue denotes copy number gain and red denotes copy number loss.

Present in 41/162 (25.3%) patients, 13q31.3 gain was most strongly associated with BL, appearing in 31/109 (28.4%) of patients 7/36 (19.4%) DLBCL, 2/11 B-NHL, NOS and the only HGBL, NOS case. All 41 cases had gain of *MIR17HG* (Figure 4.3). 16 cases with *MIR17HG* copy number gain on 13q also harboured 13qter deletion. Cases with gain-loss patterns on 13q regularly had complex patterns of gain with *MIR17HG* at the peak copy number.

13q33.1-13q34 deletion was found in 23 cases, of which 21 were BL. One DLBCL case had telomeric deletion of 13q, as did one B-NHL, NOS case. 16/21 cases had deletion of 13qter in association with copy number gain involving 13q31.3, which was a significant GISTIC2.0-identified region. Others were small abnormalities under 100kb within the GISTIC2.0 region with the exception of two cases with 8.0Mb and 5.6Mb deletions respectively. The region identified by GISTIC2.0 contained 63 genes. While the deleted region contained the *IRS2* gene which is important in

signalling, it was hypothesised that the 13qter deletions identified in these patients were primarily passenger events associated with the pattern of gain followed by telomeric loss identified in 16 cases. The gained region of 13q associated with this abnormality always involved the *MIR17HG* locus, which has been previously shown to be important in lymphoma (Scholtysik *et al.*, 2010; Schiffman *et al.*, 2011; Dal Bo *et al.*, 2015; Havelange *et al.*, 2016b).

A region of copy number gain at 11q23.3 was identified in 22 cases. In 12/22 (54.5%) cases with this abnormality this was associated with a complex 11q abnormality involving deletion of the 11q telomere (Figure 4.4). This abnormality is the hallmark of a B-NHL subtype comprised of MYC-negative Burkitt-like cases with 11q aberrations. The region spanned 36 genes, most notably including *KMT2A* which is a recurrent translocation target in multiple subtypes of leukaemia and DLBCL (Gindin *et al.*, 2015). In BLL-11q cases this gene has been shown to be repeatedly amplified beyond the copy number of the adjacent gained region (Salaverria *et al.*, 2014).

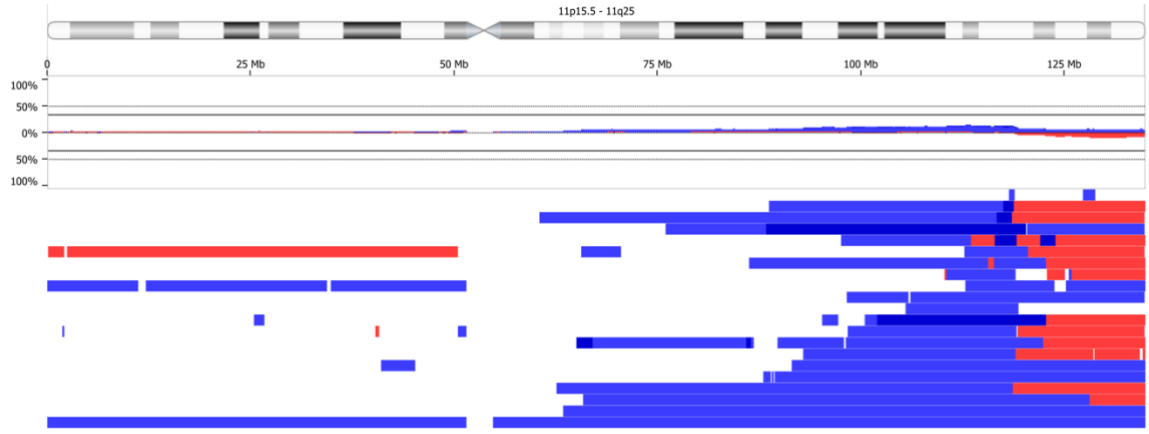


Figure 4.4 - A plot showing copy number data for chromosome 11 from 22 cases with gain of 11q23.3. Blue denotes copy number gain and red denotes copy number loss.

11q24.1-11q25 deletion was frequent in the paediatric B-NHL cohort with an equivalent q-value to 9p21.3 deletion (q=1.1x10^{-9}) and the same incidence (21/162 patients). The abnormality was primarily associated with BLL-11q cases, with 5/5 harbouring the deletion as part of the characteristic 11q aberration that defines the B-NHL subtype. Additionally, 11q24.1-11q25 deletion was seen in 12 BL cases (11.0%), of which seven had the same 11q aberration pattern as seen in BLL-11q. Other deletions were small in three cases but were contained within the GISTIC2.0-identified region, and two cases had larger telomeric deletions without an associated

region of copy number gain. The 11.2Mb region of deletion highlighted by GISTIC2.0 analysis contained 94 genes, most notably including *ETS1*, *FLI1* and *CHEK1*.

3q29 gain was observed in 17/162 paediatric cases in this cohort. The region identified by GISTIC2.0 was 2.3Mb in size and encompassed 37 genes. The region of recurrent copy number gain involved the telomeric region of the 3q arm in multiple different genomic contexts (Figure 4.5). High copy number gain/amplification was observed in four cases while three cases had copy number gain of the majority of chromosome 3. One case (21/153) harboured a chromothripsis-like event of chromosome arm 3q involving a gain of 3q29. The most frequently gained region within 3q29 included 25 genes, most notably *TFRC* and *PAK2* present in 16/17 cases. Remarkably, 6/17 cases with 3q29 gain experienced disease progression, suggesting that it may be a prognostic marker.

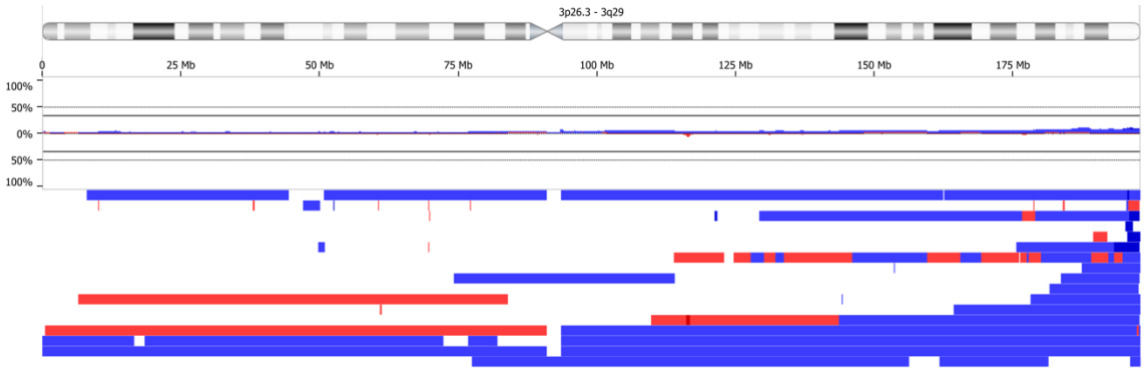


Figure 4.5 - A plot showing copy number data for chromosome 3 from 17 patients with 3q29 gain. Blue denotes copy number gain and red denotes copy number loss.

GISTIC2.0 identified a region of recurrent gain spanning 1p12-1q21.3 that was observed in 66/162 (40.7%) cases, including 49 BL patients. GISTIC2.0 did not separate a series of small copy number gains observed at the 1p side of the centromere from other frequent gains of 1q, meaning that in some samples the region of copy number gain appears to include the centromere (Figure 4.6). In reality as the centromere is not covered by the arrays the presence of an additional copy of the centromeric DNA cannot be confirmed by this method.

The recurrent gains on 1q represent the most frequently observed copy number change in BL. 21/162 (13.0%) cases had copy number gain of more than half of the

length 1q which is consistent with previous studies' findings (Lones *et al.*, 2004; Onciu *et al.*, 2006; Scholtysik *et al.*, 2010) (Figure 4.6).

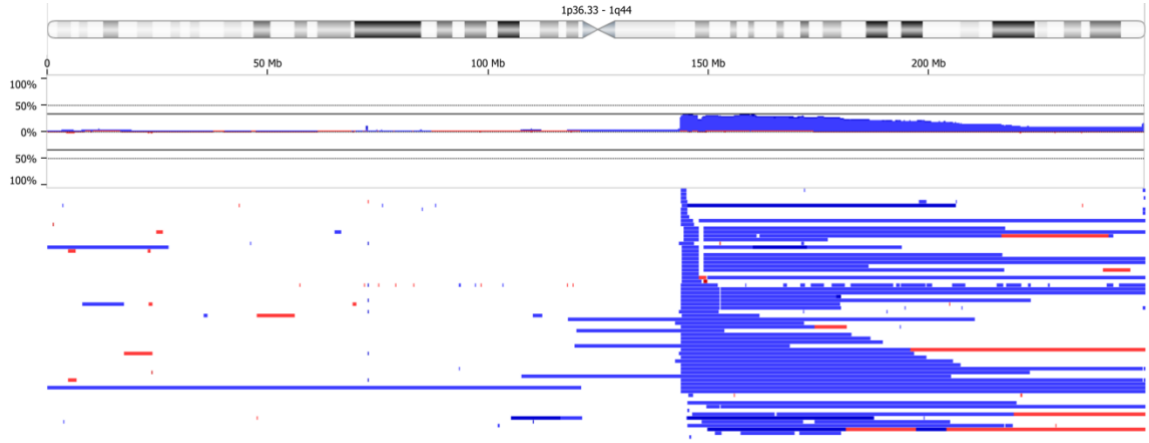


Figure 4.6 - A plot showing copy number data for 1q from 66 cases with gain of 1q. Blue denotes copy number gain and red denotes copy number loss.

Analysis of the aggregate plot for 1q copy number gain shows two separate centromeric peaks of amplification. These are not likely to be separate events but represent the fact that there are gaps in the probe coverage in this region due to the fact that 1q21.2 is repetitive. In multiple cases seen in figure 4.6 the segmentation algorithm did not bridge these two segments as the gap was larger than the defined maximum contiguous probe spacing considers and is considered too large to bridge the adjacent segments. There was nothing to suggest that these segments are separate. Interestingly, samples that were hybridised using the Affymetrix Cytoscan HD array kit did not experience this issue. This copy number array kit had fewer gaps in coverage within this cytoband and the coverage was sufficient for segmentation to be performed correctly, without over-segmentation. Cases taken from the literature that were hybridised to older Affymetrix Mapping 250K (both Nsp and Sty versions) and Affymetrix Mapping 500K array data had particularly pronounced gaps in coverage in this region. The *MCL1* gene resides within the region identified by GISTIC2.0 at 1p12-1q21.3 within the minimum common region of amplification and the gene was gained in 49/66 (74.2%) cases. The *MCL1* gene is a BCL2 family member and is highly expressed in the majority of high grade non-Hodgkin lymphomas including BL and DLBCL (Cho-Vega *et al.*, 2004). High expression of the Mcl1 protein has been shown to drive haematological malignancies and specifically *MYC*-driven lymphomagenesis in mouse models (Campbell *et al.*, 2010) and reducing expression of MCL1 is known to sensitise B-NHL cell lines to the selective BCL2 inhibitor Venetoclax (Phillips *et al.*, 2015).

18q21.2-18q21.31 copy number gain was seen in 15 cases. The 3.5Mb GISTIC2.0 region contained 16 genes and all cases had copy number gain of the genes *RAB27B* and *CCDC68*. 14 cases also had gain of *TCF4* within the region highlighted by GISTIC2.0. This region did not include the *BCL2* gene which was 5.7Mb telomeric to the region. 8/15 cases with gains in the GISTIC2.0 region also had gain of *BCL2*. The region was significantly more frequently gained in DLBCL patients than BL patients (6/36, 16.7% vs 5/109, 4.6%; p=0.03, Fisher exact test). The region identified at 14q24.3 was determined to be an artefact. 14q gain was seen in 3 cases, but the focal region was defined by recurrent small amplification segment found only in cases analysed using the Affymetrix Cytoscan HD array kit.

Gain within 2p16.1-2p15 was found in 10 cases and spanned 24 genes. The putative gene of interest here is the *REL* gene, which is part of the NFκB family of genes. Copy number gain of *REL* was seen in eight cases, with high level amplification in three cases, of which two were BL and one was DLBCL. It has previously been reported that while *REL* amplification is frequent in adult DLBCL of the GCB-like subtype, it is not seen in paediatrics. However, the one DLBCL case in this cohort with a *REL* amplification was aged 7 at diagnosis. The 3.2Mb region also included *BCL11A*, overexpression of which has been shown to play a role in lymphomagenesis and resistance to apoptosis (Gao *et al.*, 2013).

Copy number gain at 10q21.1-10q21.3 was seen in eight cases. Three cases shared a complex pattern of copy number gain between 10q21.1-10q21.3 paired with a deletion of the centre of 10q. Two cases harboured small gains in the region, but the other 6 had gains that spanned the majority of the GISTIC2.0 region. The *CDK1* gene sits within the gained region, which is vital for G1/S transition regulation.

The most significant deleted region identified by GISTIC2.0 analysis was deletion at 9p21.3 with a q-value of 1.1×10^{-9} computed by GISTIC2.0, denoting high significance. (Figure 4.7A). Deletion in this region was identified in 21/162 cases. The 0.6Mb region included three genes: *CDKN2A*, *CDKN2B* and *C9orf53* (*CDKN2A-AS1*).

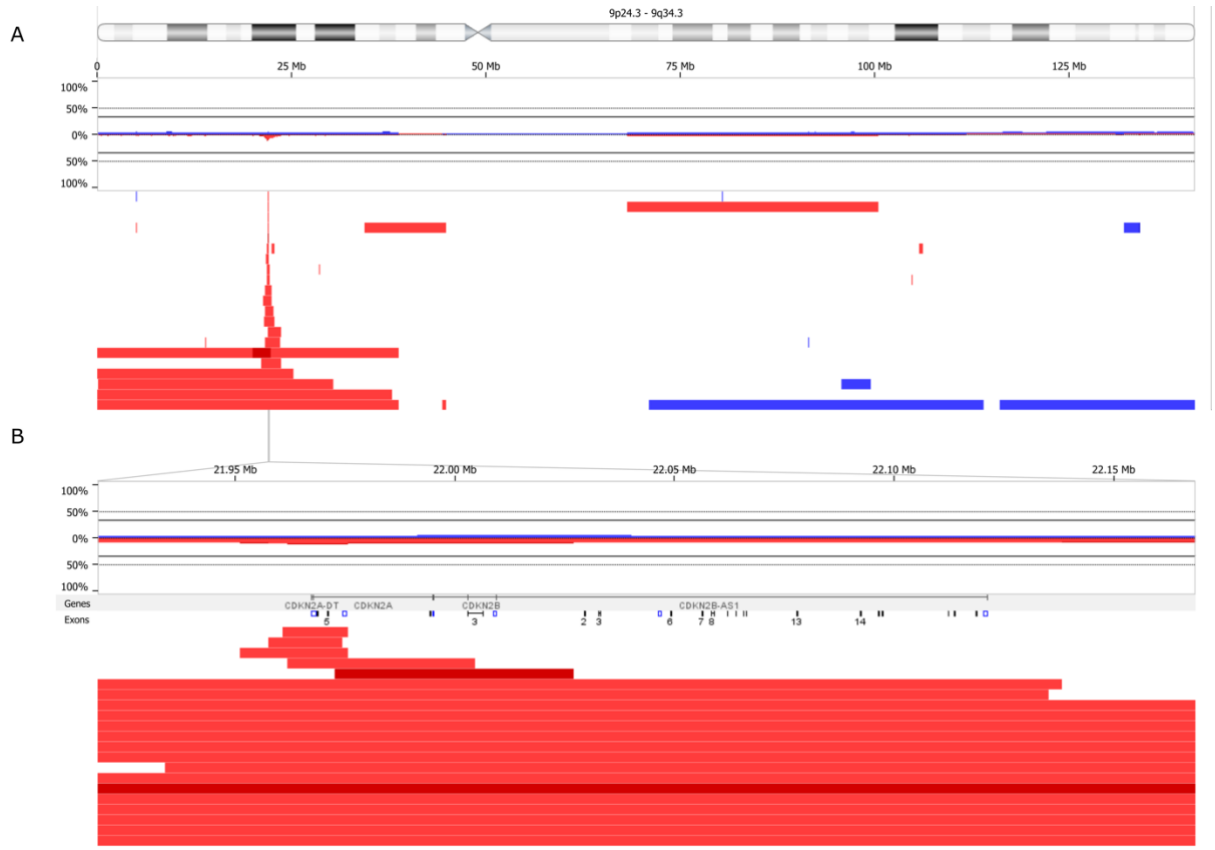


Figure 4.7 - A plot showing copy number data for (A) chromosome 9 and (B) the CDKN2A gene from 21 patients with deletion at 9p21.3. Blue denotes copy number gain and red denotes copy number loss.

In three of these cases deletions were small and only involved exon 1 on *CDKN2A* while the other 18 covered both *CDKN2A* and *CDKN2B* (Figure 4.7B). The deletions ranged from 14.9kb to 38.8Mb with a mean size of 6.0Mb. 9p21.3 deletion was present in both BL and DLBCL cases at similar incidences (BL 11.9% vs DLBCL 16.7%). *CDKN2A* codes for both $p16^{INK4a}$ and $p14^{ARF}$, the latter of which is an important inhibitor of MDM2. Deletion of the *CDKN2A* tumour suppressor gene is frequently reported in multiple cancers and is a known mechanism of p53 abrogation (Agrawal *et al.*, 2006).

A 2.5Mb significantly deleted region of 3q13.31-3q13.32 observed in 19 cases (q=7.7x10^{-7}). The majority of these cases were diagnosed with BL (12/19, 63.2%). Only one gene was situated in the focal GISTIC2 region locus, *TUSC7* (Figure 4.8).

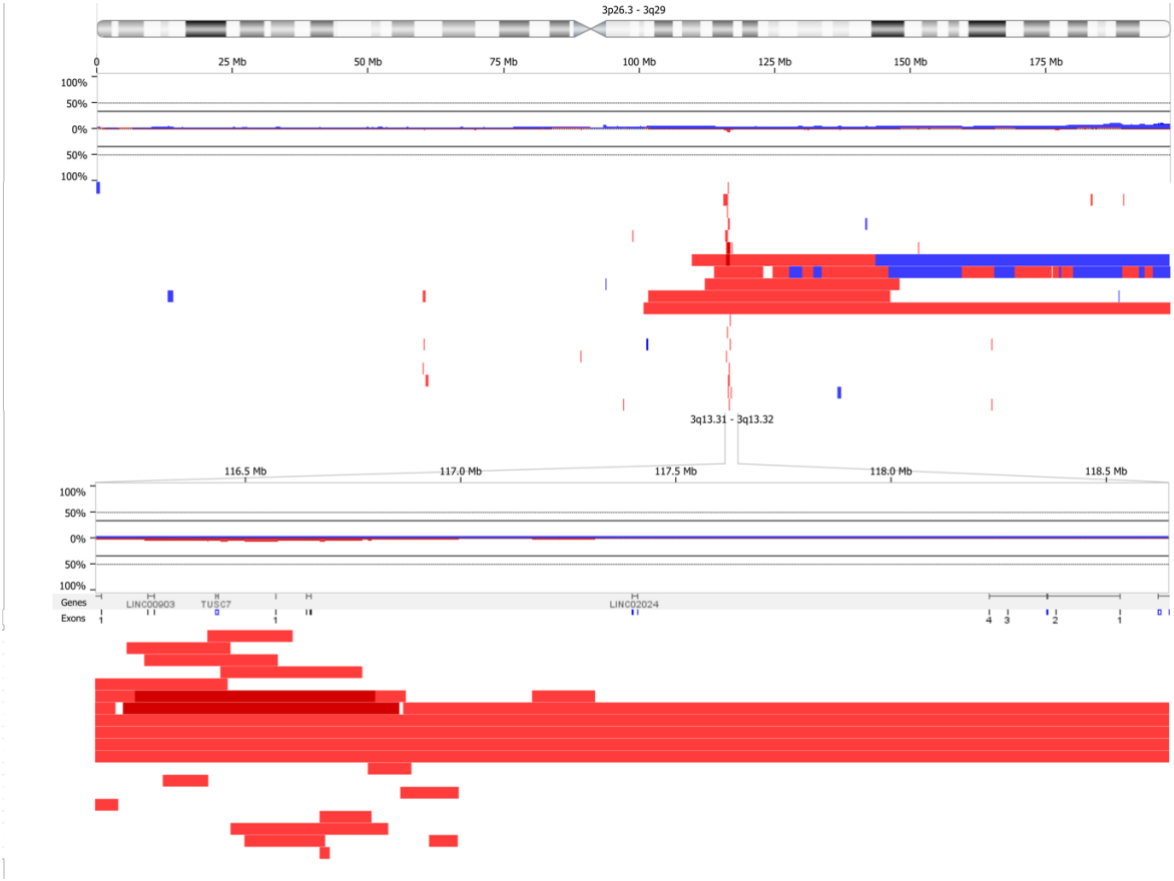


Figure 4.8 - A plot showing copy number data for (A) chromosome 3 and (B) the 3q13.31-13q31.32 GISTIC2.0 region of loss in 19 patients. Blue denotes copy number gain and red denotes copy number loss.

Deletion of the *TUSC7* gene has been identified in renal cell carcinoma (Chen *et al.*, 2003) and osteosarcoma (Kresse *et al.*, 2009) previously, and low expression is prognostic in ovarian cancer (Ntougkos *et al.*, 2005). Two BL cases harboured biallelic deletion of the *TUSC7* gene at diagnosis.

Loss of 2p22.2-2p16.3 was observed in 15 cases. The region was 12.7Mb in size and spanned 75 genes, most notably the mismatch repair family genes *MSH2* and *MSH6* as well as the F-box family gene *FBXO11*, which has a role in BCL6 degradation (Duan *et al.*, 2012). Two BL cases had smaller, focal biallelic deletions of the overlapping region between the 3' ends of the *MSH6* and *FBXO11* genes. Deletions within the 12.7Mb region were identified in 11 BL cases, 3 DLBCLs and one B-NHL, NOS. The deletions identified in the B-NHL, NOS case and one BL case (4/830 and 1/49 respectively) were part of a larger complex chromosomal abnormality spanning the rest of 2p. Both complex abnormalities also harboured a copy number gain of *REL*, also found on 2p.

6q13-6q22.1 deletion was detected in 44 cases. The 42.7Mb region covered the majority of the 6q, while only 15 cases had deletions that spanned the more than half of the arm. 6q harbours several genes that have previously been associated with lymphoma and deletion of the 6q arm has been reported frequently in the past. Deletions in *GRIK2* (Sinclair *et al.*, 2004), *PRDM1* (Pasqualucci *et al.*, 2006) and *BACH2* (Swaminathan *et al.*, 2013) have all been seen in haematological malignancies, with the latter two reported in B cell lymphoma. *PRDM1* deletions were only seen in cases without disease progression (0/18, 0.0% vs 18/119, 15.1%, p=0.13, Fisher exact test). The incidence of deletions within the 6q GISTIC2.0 focal region were consistently similar between BL and DLBCL subgroups (23/109, 21.1% and 8/36, 22.2% respectively).

19p13.3-19p13.2 deletion involved a small 0.6Mb region with 17 genes. Eight cases were found to have this deletion, of which six were DLBCL cases (75%). The other two cases harbouring the deletion were a BL case and a B-NHL, NOS case. The *TNFSF14, TNFSF9* and *TNFSF7* (*CD70*) genes were lost in eight, seven and eight cases respectively. A study has reported loss of these genes in DLBCL and BL previously (Scholtysik *et al.*, 2012). Two cases harboured homozygous deletions of the focal region involving *TNFSF7* and *TNFSF14*. One such deletion was standalone with no other abnormalities while the other occurred within a larger 7.4Mb region of deletion. Of the six hemizygous-only deletions, one spanned the entirety of chromosome 19, three were small and ranged from 500kb and 990kb in size while remaining focal to the TNFSF family genes, and two were more complex with immediately adjacent regions of copy number gain.

A 43.8Mb region spanning cytobands 4q12 to 4q22.3 detected GISTIC2.0 analysis. The region is adjacent to the centromere on 4q and deletions in this locus was identified in 32 cases, however most were small focal deletions of *UGT2B17*. The *UGT2B17* gene has been shown in prostate cancer (Gallagher *et al.*, 2007) but the relevance of the gene in lymphoma is unclear. Larger deletions were identified in ten cases with a 3.5Mb minimal common region involving 26 genes, none of which had previously been associated with lymphoma pathogenesis.

8p telomeric deletion was observed in nine cases, four of which spanned the rest of the arm. The 8p23.3-8p23.2 GISTIC2.0 region was infrequently observed in both BL

(5/109, 4.6%) and DLBCL (3/36, 8.3%) as well as one B-NHL, NOS. The region includes the large gene *CSMD1*, which is known to inhibit classical complement pathway activation and has been proposed as a tumour suppressor gene in breast cancer (Escudero-Esparza *et al.*, 2016). Low expression and deletion were shown to be associated with a poorer overall survival in breast cancer. Although rare, deletions of 8p23.3-8p23.2 were found in two BL cases who underwent disease progression. There were too few cases to perform conventional survival analysis, however.

GISTIC2.0 identified a recurrent region of deletion between 1q25.1-1q44. 30 cases had deletions within these cytobands, but the majority were small. Larger deletions were observed in eight cases with a mean size 30.0Mb and ranged between 1.6Mb and 64.7Mb in size. Six of these deletions were observed as part of complex abnormalities involving centromeric gain followed by telomeric loss similar to those seen on 11q and 13q.

4.3.5 Recurrent regions of copy number neutral loss of heterozygosity (CNN-LOH)

Analysis of regions of CNN-LOH within the cohort revealed potential sites of recurrent mutation. CNN-LOH is not well characterised in B-NHL, with relatively few studies investigating recurrent CNN-LOH. A copy number investigation of 148 DLBCL identified no recurrent CNN-LOH events in their cohort (Scholtysik *et al.*, 2015). In contrast, a study of 39 BL cases found CNN-LOH in nine cases, with 17p, 17q and 6p CNN-LOH being recurrent (Scholtysik *et al.*, 2010). Due to the 10Mb filter for regions of copy number neutral allelic events the results of focal and broad analysis from GISTIC2.0 were very similar, particularly on smaller chromosome arms.

Recurrent CNN-LOH of 1p36.33-1p34.3 was identified as the most significant region of CNN-LOH in this paediatric B-NHL cohort with a q-value of 4.0×10^{-7}. The region was large at 35.9Mb and contained 553 genes. 9/11 cases with CNN-LOH in this region were BL. The abnormalities began at the 1p telomere in all cases. The most notable of these genes was *ID3* which is frequently mutated in up to 58.5% of BL cases but is not associated with DLBCL (Schmitz *et al.*, 2012). The *TCF3-ID3* axis is known to be important in BL, with 70% of BL patients in one study having a mutation

in one or both of these genes. Also, in this region was the *TP73* gene, a member of the p53-like gene family. *TP73* plays a role in DNA-damage mediated apoptotic signalling and is considered a tumour suppressor in medulloblastoma (Castellino *et al.*, 2007).

17q CNN-LOH was observed frequently in our cohort, with 20 cases having at least one region of CNN-LOH larger than 10Mb. Two focal regions were identified by GISTIC2.0 analysis. The first, and most significant (q=6.9x10^{-4}) spanned a 10Mb region from 17q22-17q24.1. CNN-LOH in this region was found in 18 cases and contained 127 genes, notably including *GNA13* (Figure 4.9).

Figure 4.9 - a plot showing CNN-LOH calls on 17q in 19 patients with 17q22-17q24.1 CNN-LOH. Yellow denotes CNN-LOH.

GNA13 is known to be mutated recurrently in BL cases. Frequency of mutation ranges in the literature from 17.1% to approximately 30% (Love *et al.*, 2012; Schmitz *et al.*, 2012). The *GNA13* gene codes for a guanine nucleotide-binding G protein and abrogating mutations are thought to be an oncogenic driver event in germinal centre B-like lymphomas and BL (Love *et al.*, 2012). Abnormalities in *GNA13* are usually inactivating mutations or deletion. While deletion was seen in 5/162 (3.1%) cases, and none were focal to the GNA13 gene, CNN-LOH involving the gene occurred in 19/162 (11.7%) cases. 14/19 cases with abnormality in this region were diagnosed as BL, with 4 DLBCL and one B-NHL, NOS case.

The second region on 17q identified by GISTIC2.0 involved part of the chromosome nearer the centromere between 17q11.2-17q21.32. The region was larger at 18.7Mb

and harboured 444 genes (Figure 4.10). This region was less frequently observed than the other region on 17q, being seen in only 15/162 cases (9.3%, *q*=0.053). In a similar manner to the 17q22-17q24.1 region, the majority of cases were BL (13/15) while it was observed in one DLBCL and one B-NHL, NOS.

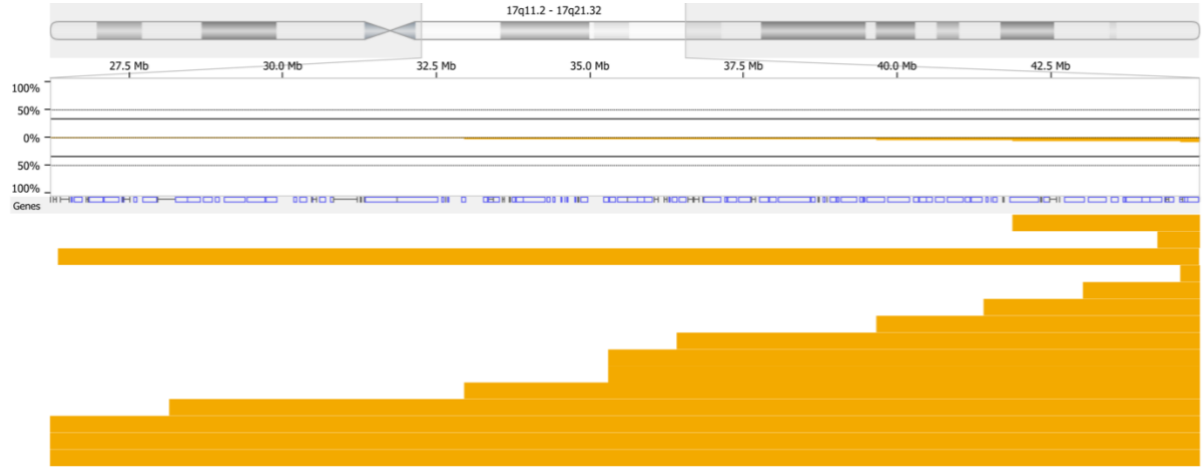


Figure 4.10 - A plot showing CNN-LOH calls on 17q in 15 patients with 17q11.2-17q21.32 CNN-LOH. Yellow denotes CNN-LOH.

CNN-LOH involving 1q25.3-1q44 was seen infrequently in nine patients. Chromosome 1q is regularly altered in paediatric B-NHL, with both copy number gains and losses also being reported by GISTIC2.0. CNN-LOH of the arm was less frequent than gain or loss and no region of 1q was aberrant in all nine cases. More than half of 1q underwent CNN-LOH in six of the cases and all of these aberrations involved the 1q telomere. The other three cases had smaller aberrations 11.4Mb, 19.4Mb and 21.1Mb in size respectively that did not overlap in any two cases. The GISTIC2.0 region on 1q contained 521 genes, of which 15 were identified in 'Pathways in Cancer' within KEGG Mapper (Kanehisa *et al.*, 2019). Two genes of the WNT family were found within this region, *WNT9A* and *WNT3A*. *MDM4* was also identified in the region, which has a similar function to *MDM2* in chelation and ubiquitination of p53 protein. Previous studies in BL have suggested that up regulation of *MDM4* may functionally inhibit p53 in cases without *TP53* or *MDM2* abnormalities (Leventaki *et al.*, 2012).

A large region of CNN-LOH on chromosome 9p was detected spanning 21.8Mb which also encompassed the C*DKN2A/B* locus, which could suggest the presence of driver mutations in this locus. The region encompassed 99 genes in total and was observed in nine cases. The locus also included the *JAK2* gene which frequently

harbours a V617F mutation in haematological malignancies, however rarely in non-Hodgkin lymphomas (Lee *et al.*, 2006).

CNN-LOH at 6p25.3-6q11.1 was seen in 13 cases. 11 cases showed CNN-LOH across 6p only, with two cases having 6q CNN-LOH and whole chromosome 6 CNN-LOH respectively. The region most frequently observed to undergo CNN-LOH involved 243 genes in a 21.7Mb stretch of genome. Three genes associated with EBV infection are found in the region, of which two were HLA genes (*HLA-F*, *HLA-G*) as well as the transcription factor *E2F3*.

15q CNN-LOH was seen in 6/162 cases with three BL, one DLBCL, one BLL-11q and one B-NHL, NOS cases affected. The minimal common region of abnormality was 6.1Mb in size and covered 79 genes while the wide peak identified by GISTIC2.0 was significantly larger at 42.6Mb with 383 genes. Of the genes in the common region, *B2M* was considered a candidate gene for mutation due to a known role in evading immune recognition in DLBCL when combined with *CD58* inactivation (Challa-Malladi *et al.*, 2011).

17p13.3-17p11.2 CNN-LOH affected 14 cases that were primarily diagnosed BL (12/14, 85.7%) (Figure 4.11). The region identified spanned the majority of the p arm of chromosome 17 and involved 314 genes. 210 of these genes lost heterozygosity in all 14 cases affected and included *TP53*.

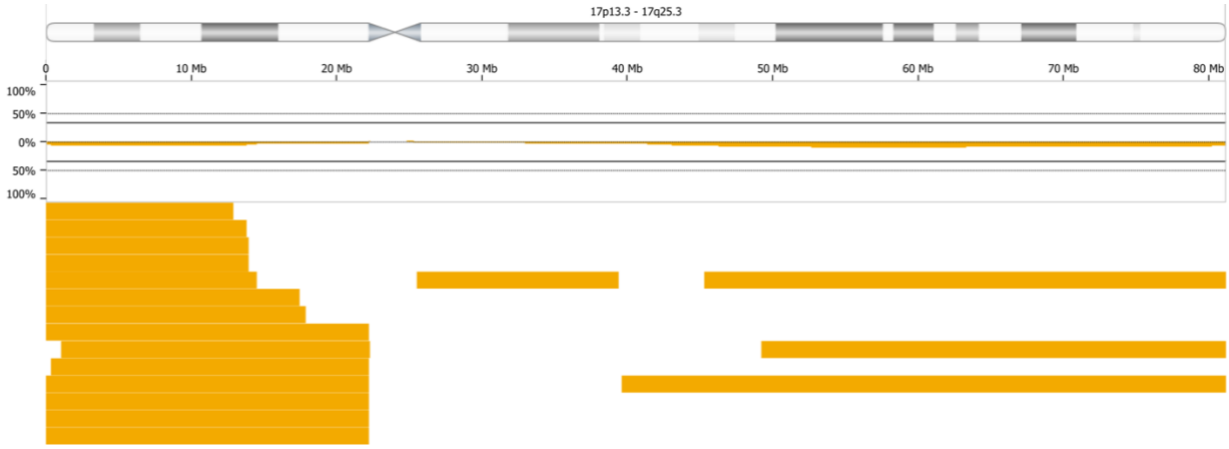


Figure 4.11 - A plot showing CNN-LOH calls on chromosome arm 17p in 14 cases. Yellow denotes CNN-LOH.

TP53 is commonly perturbed in many human cancers and is reported to be mutated in approximately 20% of BL cases (Love *et al.*, 2012; Schmitz *et al.*, 2012; Havelange *et al.*, 2016b). *TP53* aberration is a prognostic marker in several tumour types, including another B cell malignancy, chronic lymphocytic leukaemia (CLL) (Li *et al.*, 2019) and is known to have numerous important roles in cell proliferation and evasion of senescence, DNA damage response and resistance to apoptosis.

4.3.6 GISTIC2.0 Broad Analysis to detect large scale copy number abnormalities

Broad analysis identified gain of seven chromosome arms, loss of two and CNN-LOH of four with gain of 1q, 7q and loss of 17p as the most common regions of CNAs occurring in 13% of cases each (21/162). Most regions detected by broad analysis aligned with regions detected by focal analysis, while 7p, 7q, 12p and 12q large gains were not picked up by GISTIC2.0. Gain of chromosome 7 was common in both BL and DLBCL but BL cases tended to have gain of 7q only. 7q gain was most common as 21/162 cases had gain of more than half of 7q. A minimal common region of gain seen in 30/162 (18.5%) cases at 7q21.3-7q22.1 containing 67 genes. The region contained *MCM7*, a component of DNA replication machinery and correct assembly of replication forks, amplification and overexpression of which is prognostic in prostate cancer (Ren *et al.*, 2006). The region also harbours the *MIR106B* microRNA cluster and functional paralogue of *MIR17HG*.

Gains on chromosome 12 were most common in DLBCL (12p gain 10/36 vs 4/109, p=0.0001; 12q gain 6/36 vs 2/109, p=0.003, Fisher exact test). No peak or high copy number amplification was observed on chromosome 12. 16 cases in the cohort had gain of *MDM2* at 12q15 which is a common abnormality in cancer. High copy number amplification involving multiple additional copies of *MDM2* is associated with upregulation of MDM2 protein, which ubiquitinates p53 protein and targets it for degradation. p53 is abrogated in many tumours where the *TP53* gene is wildtype by high expression of *MDM2* (Olivier *et al.*, 2010).

4.3.7 Comparison of aberrations in BL and DLBCL

Abnormalities in paediatric B-NHL were most frequent on chromosome arms 1q, 6q, 7q and 13q, with recurrent aberrations of 11q and 17p (Figure 4.12). Broadly, BL-

associated abnormalities were consistent with those found in the literature previously. 1q gain was the most frequently observed abnormality in BL, with 49/109 (45.0%) harbouring copy number gain within the region identified by GISTIC2.0, compared to 9/24 (37.5%) and 10/39 (25.6%) reported in previous copy number array studies (Scholtysik *et al.*, 2010; Havelange *et al.*, 2016b). The higher reported incidence in this study is likely a result of the large proportion of the cohort analysed using higher resolution Affymetrix Cytoscan HD arrays, allowing for confident calling of smaller aberrations.

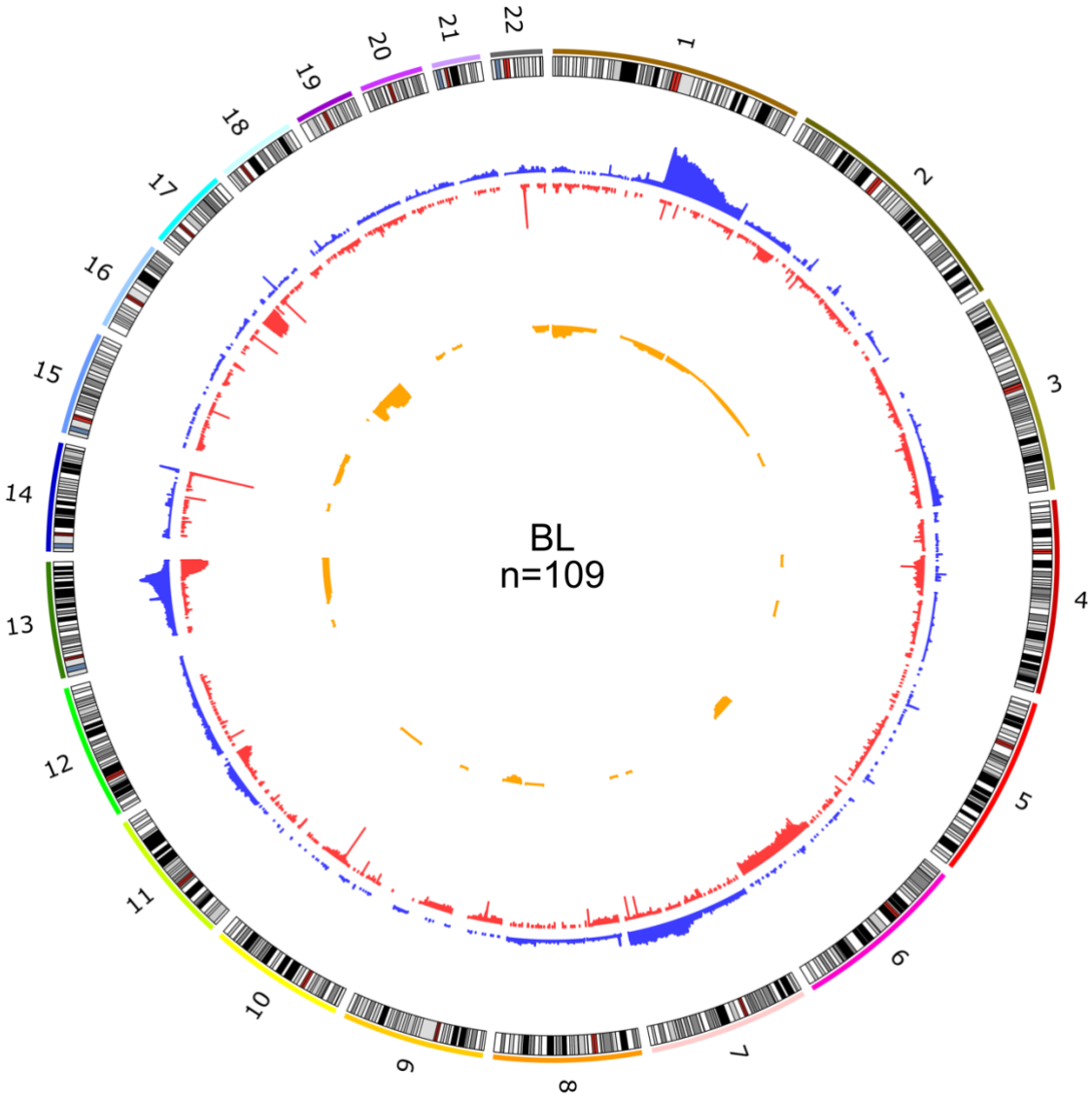


Figure 4.12 - A Circos plot showing copy number profiles for 109 paediatric BL patients. Blue denotes copy number gain, red denotes copy number loss and yellow denotes CNN-LOH.

Considering copy number abnormalities not highlighted by GISTIC2.0, multiple regions were identified that were enriched in one disease subtype compared to the others. Gain of chromosome 7 has been repeatedly reported as recurrent and potentially prognostic in BL (Poirel *et al.*, 2009; Nelson *et al.*, 2010; Scholtysik *et al.*, 2010). 11/109 (10.1%) BL patients in our cohort had large gains on 7q which correlates with previously reported incidences of 6/39 (15.4%) (Scholtysik *et al.*, 2010) and 36/238 (15.1%) (Poirel *et al.*, 2009) (Figure 4.12). The most commonly gained region was 7q21.3-7q22.1 and was seen in 14.7% (16/109) of BL tumours investigated. The region encompassed the *MIR106B* miRNA cluster, which is paralogous to the recurrently amplified *MIR17HG* gene. 13q abnormalities were common in the BL subgroup with recurrent *MIR17HG* amplification (31/109, 28.4%) and telomeric deletion (21/109, 19.3%). Of these, 16 patients had both amplification of *MIR17HG* at 13q31.3 and subsequent telomeric deletion.

Overall, the genome of DLBCL was more complex than that of BL (Figure 4.12, Figure 4.13), with significantly more of the genome altered (BL 7% vs DLBCL 15%, $p<0.001$, student t-test) and a higher mean number of CNAs per sample (24 BL vs 31 DLBCL, $p=0.040$, student t-test)(Figure 4.13). The mean number of regions of CNN-LOH, however, were similar (1.7 BL vs 2.3 DLBCL, $p=0.610$, student t-test).

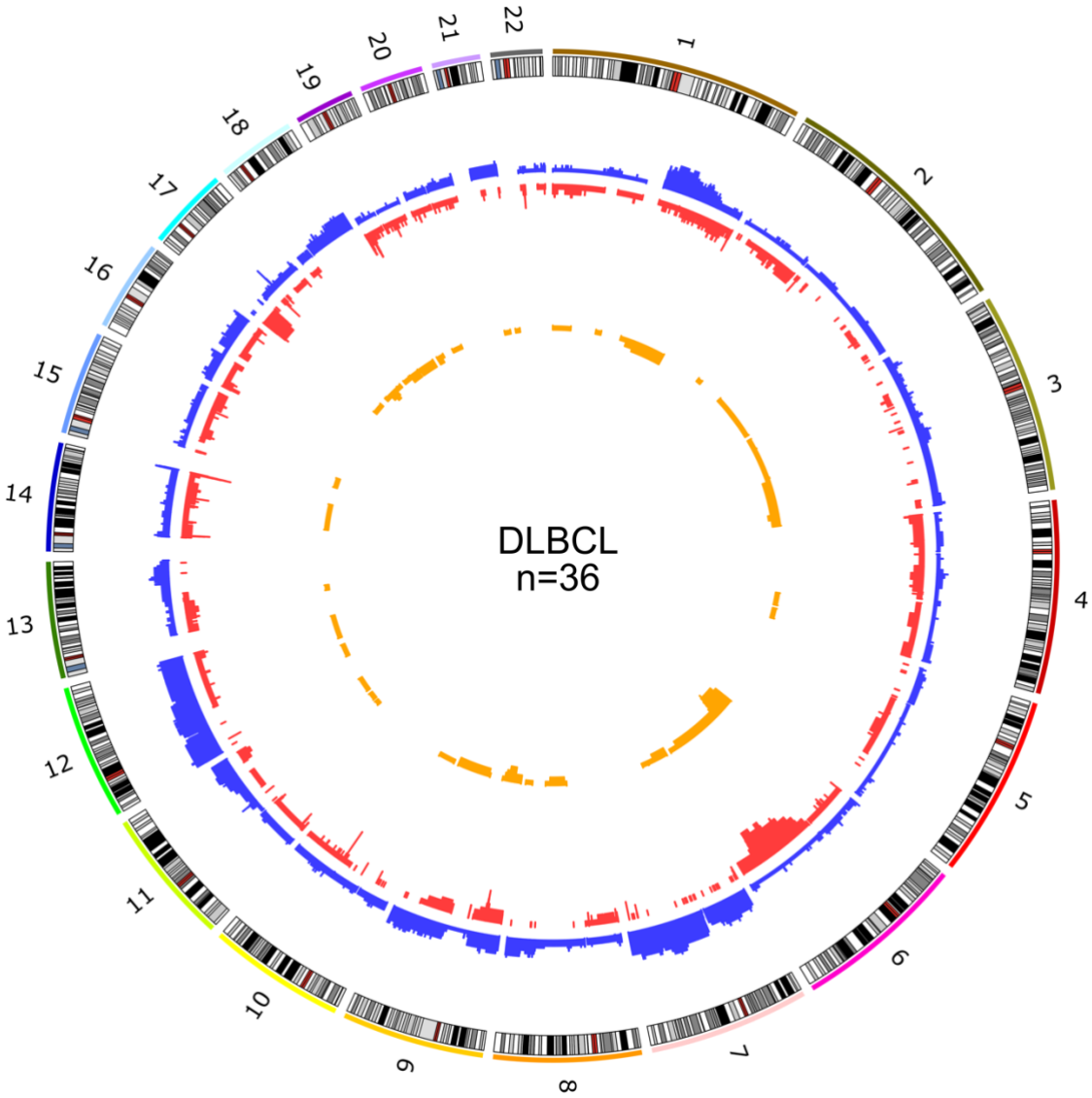


Figure 4.13 - A Circos plot showing copy number profiles for 36 paediatric DLBCL patients. Blue denotes copy number gain, red denotes copy number loss and yellow denotes CNN-LOH.

In BL, recurrent abnormalities were largely restricted to six chromosomal arms (1q, 6q, 7q, 11q, 13q and 17p), whereas most chromosomes were altered in at least one DLBCL case. Of note, gain of 1q was only observed in DLBCL cases with an *IG-MYC* translocation (3/6, 50.0% *IG-MYC* positive vs 0/15, 0.0% *IG-MYC* negative, p=0.015, Fisher exact test).

Several abnormalities were detected at similar frequency in both BL and DLBCL, for example 17p loss (14/109, 12.8%% BL vs 6/36, 16.7% DLBCL, p=0.582, Fisher exact test), however differences were observed between subtypes. Deletion of 13q33.1-13qter was more frequent in BL compared to DLBCL (21/109, 19.3% BL vs

1/36, 2.8% DLBCL, *p*=0.015, Fisher exact test), whilst gain of the *MYC* locus on chromosome 8q24.21 was more common in the DLBCL cohort (4/109, 3.7% BL vs 4/36, 11.1% DLBCL, *p*=0.105, Fisher exact test), as well as gain of chromosome 7 (3/109, 2.8% BL vs 7/36, 19.4% DLBCL, *p*=0.002, Fisher exact test). In BL, deletions of 6q were large, whereas in DLBCL these were more focal, with the minimal region of loss defined as 6q23.2-6q24.2 (11/109, 10.1% BL vs 8/36, 22.2% DLBCL, *p*=0.085, Fisher exact test), although this comparison was not statistically significant.

4.3.8 Burkitt-like lymphoma with 11q aberration (BLL-11q)

In total five cases were diagnosed as BLL-11q, a recently characterised subset of B-NHL cases defined by Burkitt-like morphology with the absence of a *MYC* rearrangement, and an 11q abnormality involving copy number gain adjacent to 11qter loss (Salaverria *et al.*, 2014; Gonzalez-Farre *et al.*, 2019). The BLL-11q cases shared a common breakpoint within a 0.6 Mb region on 11q23.3 spanning 118,605,127-119,240,071 bp involving 31 genes including *CBL*, *FOXR1*, *BCL9L*, *CXCR5*, and *DDX6*. In comparison the breakpoint of seven BL cases with a similar pattern of 11q abnormalities, were less focal and spanned a much larger 29.2 Mb region spanning 11q22.1-11q24.3 from 98,895,212-128,124,434 bp, encompassing 322 genes. Analysis of the genomic profiles revealed that 4/5 (80%) of the BLL-11q cases had loss of 6q while 1/7 (14%) BL patients with 11q abnormalities had 6q loss (*p*=0.072, Fisher exact test). The 13.2 Mb minimal region of 6q loss in BLL-11q patients included *MAP3K7*, *BACH2* and 60 other genes. In BL, 1q gain and *MIR17HG* amplification were the most frequent abnormalities, while in contrast no BLL-11q cases had either abnormality. The copy number profiles of BLL-11q cases were otherwise quiet with no recurrent abnormalities (Figure 4.14).

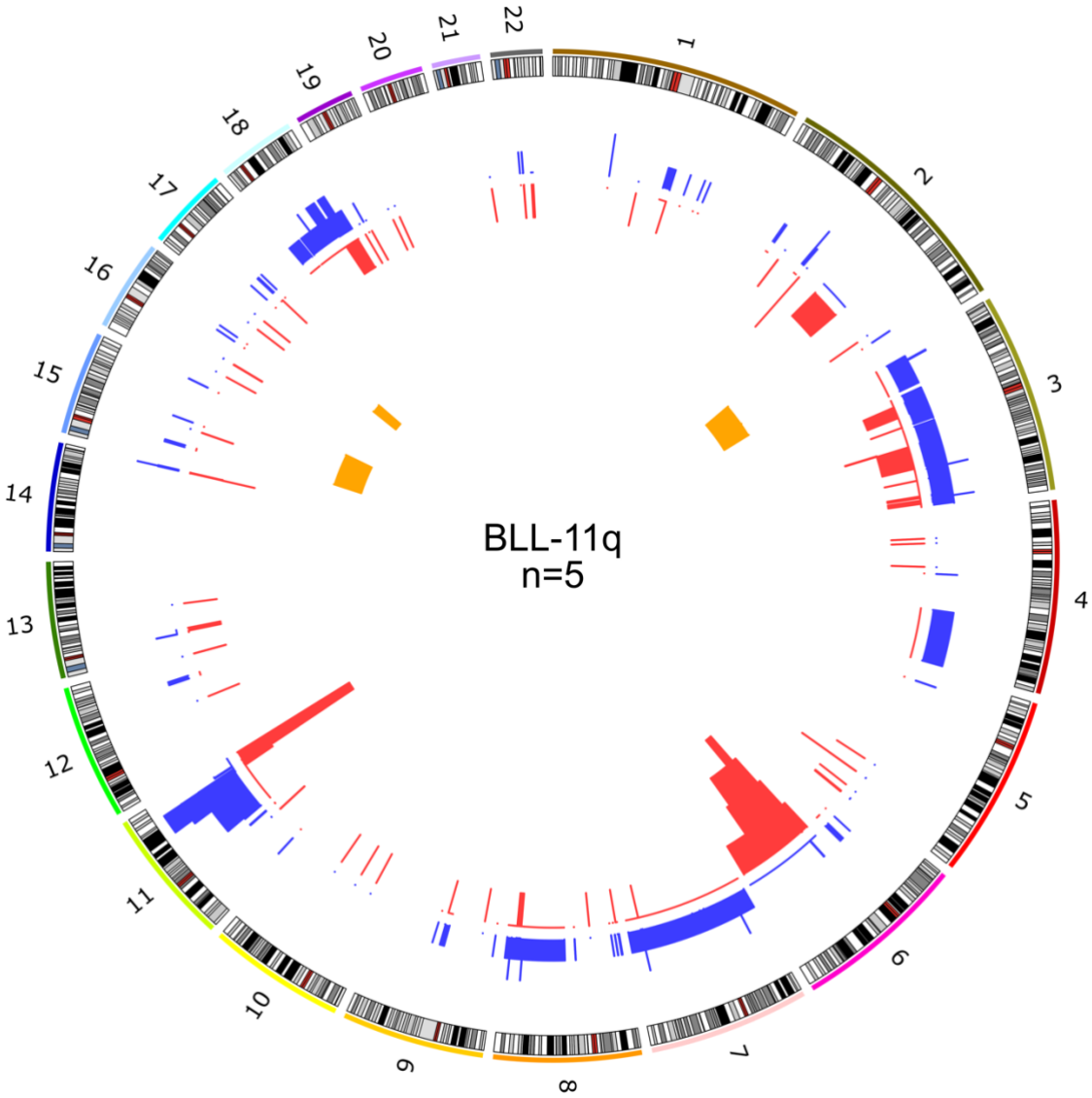


Figure 4.14 - A Circos plot showing copy number profiles for 5 patients diagnosed with BLL-11q. Blue denotes copy number gain, red denotes copy number loss and yellow denotes CNN-LOH.

4.3.9 Univariate Survival Analysis

Survival estimates at 3 years for Time to Progression and Overall Survival were 85.2% (CI 95% 79.2-91.8) and 85.1% (CI 95% 79.4-91.2) respectively. This is representative of the survival rates reported in the FAB/LMB96 trial, although slightly lower (Patte *et al.*, 2007). The slightly lower estimates for TTP and OS may be explained by treatment protocols varying between studies. A subset of patients from both Scholtysik *et al.* (Scholtysik et al., 2010; Scholtysik et al., 2015) studies are described as being treated on “AML-like therapy” and many of the patients

investigated in this analysis, both from the literature and the CCLG cohort, pre-date the FAB/LMB96 protocols being adopted as the gold standard therapy.

Survival analysis by Cox proportional hazards method was performed to understand which copy number abnormalities detected by SNP array were associated with superior or inferior overall survival or time to progression. Univariate factors were assessed individually to determine which factors should be included in multivariate analysis and modelling. All GISTIC2.0-identified focal and broad regions of copy number gain, loss and CNN-LOH were included in the analysis. Clinical parameters were also investigated including bone marrow and central nervous system involvement, diagnosis, age at diagnosis, *MYC* rearrangement status, disease stage, treatment protocol, gender and also number of copy number alterations and genome percentage of the genome changed.

Univariate analysis with TTP and OS outcome data highlighted one region of copy number gain and two regions of CNN-LOH that were significantly associated with poorer outcome (Table 4.4).

Variable	Time to progression (TTP)		Overall survival (OS)		
	HR (95% CI)	*p* value	HR (95% CI)	*p* value	No. cases with abnormality (%)
Univariate analysis					
3q29 gain	6.2 (2.3-17)	< 0.001	4.7 (1.9-11)	< 0.001	17 (10.5%)
17p13.3-17p11.2 CNN-LOH	5.6 (2-16)	0.001	3.1 (1.1-9.2)	0.038	14 (8.6%)
17q11.2-17q21.32 CNN-LOH	4.5 (1.5-14)	0.008	4.6 (1.8-12)	0.002	15 (9.3%)
17q22-17q24.1 CNN-LOH	3.4 (1.1-10)	0.031	3.4 (1.3-8.5)	0.011	18 (11.1%)
No. CNA	1 (1-1)	0.026	1 (0.99-1)	0.340	NA
Age at Diagnosis	0.89 (0.79-1)	0.057	0.94 (0.85-1)	0.220	NA

Table 4.4 - Univariate survival analysis results by Cox-proportional hazards method

3q29 gain was the third most significant region of copy number gain detected by GISTIC2.0 and was shown by univariate analysis to be associated with an inferior time to progression. 3q29 gain occurred in 17 cases throughout the cohort of which six (6/119, 5.0%) were cured by frontline therapy and another six (6/18, 33.3%) underwent disease progression (with 5 having insufficient outcome data). 3q29 gain was the most significant factor in univariate analysis for overall survival (Figure 4.15A-B).

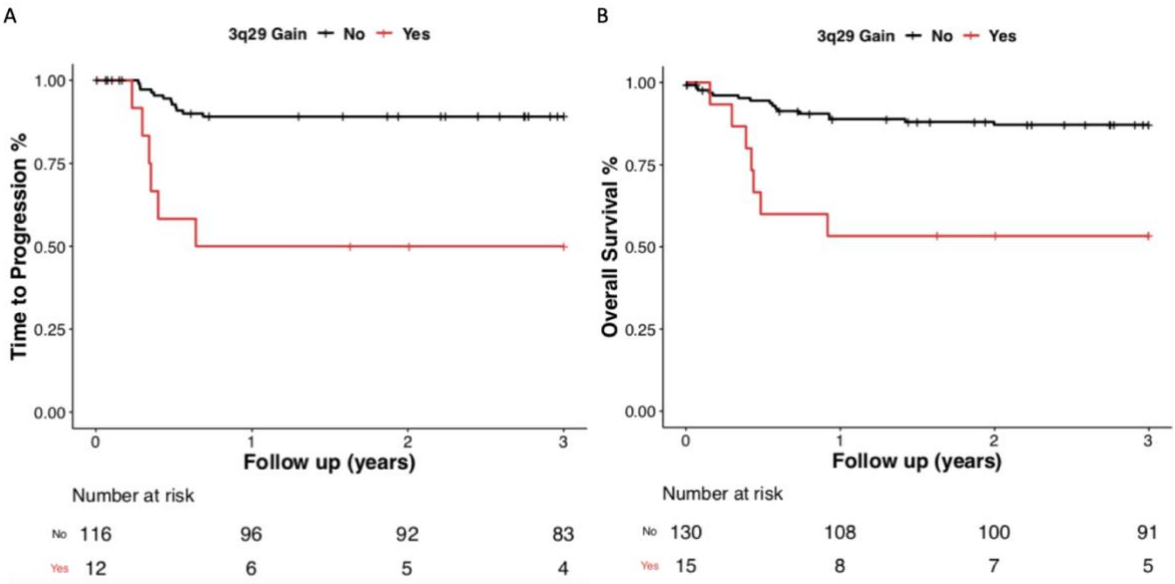


Figure 4.15 - Kaplan Meier plots demonstrating the risk associated with 3q29 gain by (A) Time to Progression ($p<0.001$) and (B) Overall survival ($p<0.001$) analyses.

The focal region of 17p CNN-LOH was shown to be associated with a poorer time to progression (HR 5.6 95% CI 2-16, p=0.0011, Cox proportional hazard method) (Figure 4.16A-B). Broad 17p CNN-LOH was also significant by the same analysis (HR 5.2 95% CI 1.5-18, p=0.0097, Cox proportional hazard method). CNN-LOH of the region was present in a higher proportion of cases who underwent relapse/refractory disease (5/18, 27.8% vs 8/119 6.7%, Fisher exact test). All aberrations in all cases involved the *TP53* gene locus which suggests that it may play an important role in outcome and tumour resistance to therapy. Further investigation of the *TP53* locus in paediatric B-NHL is performed in Chapter 5.

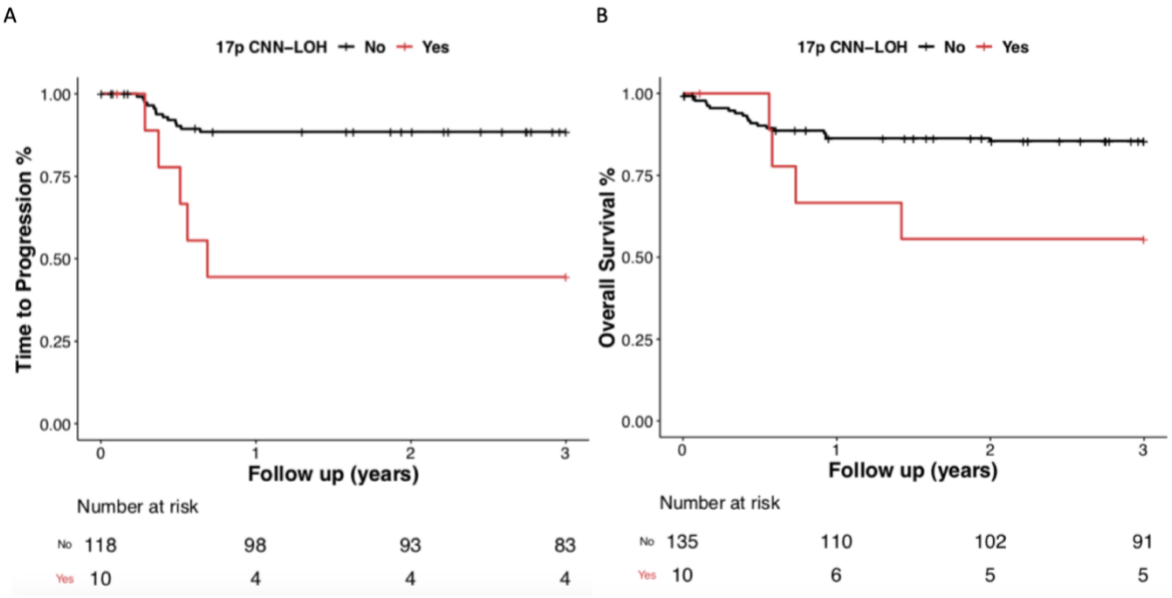


Figure 4.16 - Kaplan Meier plots demonstrating the risk associated with 17p CNN-LOH by (A) Time to Progression ($p<0.001$, Log rank test) and (B) Overall survival ($p=0.029$, Log rank test) analyses.

Both regions of 17q CNN-LOH were significant markers of inferior time to progression. The centromeric region spanning 17q11.2-17q21.32 CNN-LOH was most significant with the highest hazard ratio (HR 4.5 95% CI 1.5-14, p=0.008, Cox proportional hazard method) (Figure 4.17A-B). The alternative 17q CNN-LOH region at 17q22-17q24.1 had a lower hazard ratio than the region focal to the centromere (HR 3.4 95% CI 1.1-10, p=0.031, Cox proportional hazard method) but was still a significant survival factor. This region contained the *GNA13* gene which may suggest that mutations may confer an inferior prognosis and higher risk of disease progression. This warrants further investigation of mutation data with a focus this region to determine whether *GNA13* mutations are associated with disease progression or outcome. All 15 cases with CNN-LOH at 17q11.2-17q21.32 also had CNN-LOH at 17q22-17q24.1 and all 19 cases with CNN-LOH on 17q had aberrations involving *GNA13*.

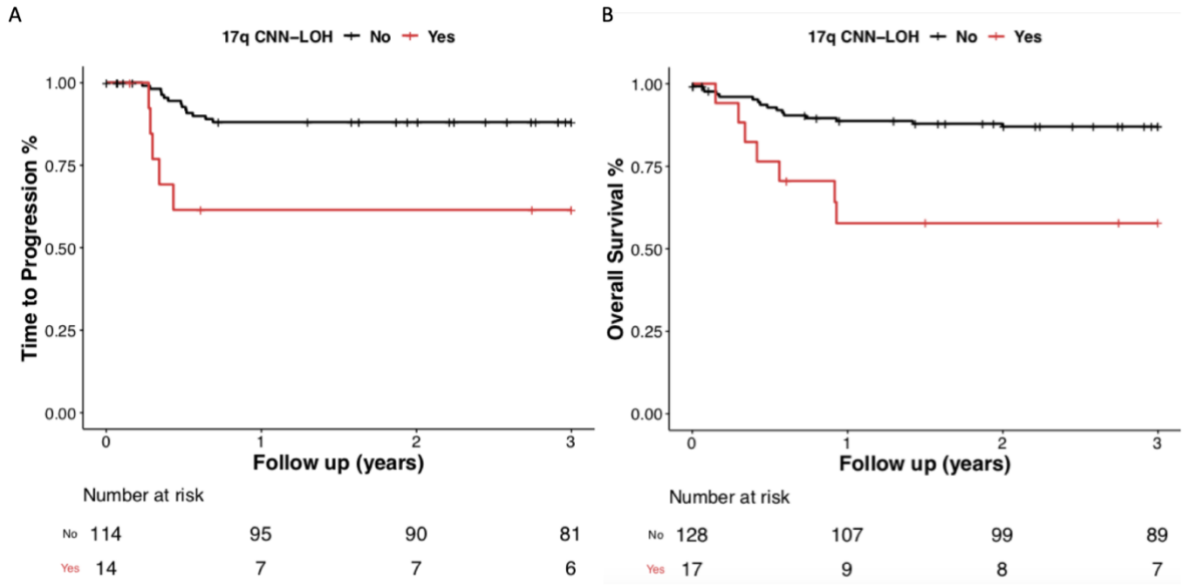


Figure 4.17 - Kaplan Meier plots demonstrating the risk associated with 17q CNN-LOH by (A) Time to Progression (p=0.003, Log rank test) and (B) Overall survival (p=0.001, Log rank test) analyses.

4.3.10 Multivariate Prognostic Modelling

Multivariate survival modelling was performed using a forward selection method. Factors included in the model had univariate Cox proportional hazards *p* values below 0.1. Several factors were assumed to represent the same genomic aberration, most notably broad 17p CNN-LOH and focal 17p CNN-LOH. In cases where this occurred only the focal region was included as this better represented the minimal common region and nature of the abnormality in the setting of B-NHL. The first factor included in the model was the most significant (3q29 gain) which made up Fit1, followed by the first round of selection in which all other univariate factors where $p<0.1$ were assessed in the model alongside 3q29 gain separately. A likelihood ratio test was performed on each model as part of the first forward selection to determine the strength of the model compared to Fit1. Five factors were tested in the first forward selection: 17p13.3-17p11.2 CNN-LOH, 17q11.2-17q21.32 CNN-LOH, total number of copy number abnormalities, 17q22-17q24.1 CNN-LOH and patient age. Three models were significantly stronger than Fit1: 17p13.3-17p11.2 CNN-LOH, 17q11.2-17q21.32 CNN-LOH and 17q22-17q24.1 CNN-LOH. The most significant was the addition of the focal region of CNN-LOH on 17p to 3q29 gain ($p=0.012$, Likelihood ratio test) and this model was taken forward to second forward selection as Fit2.

The second forward selection compared Fit2 against itself alongside one of four additional factors. Only one model was significantly stronger than Fit2, which included 17q11.2-17q21.32 CNN-LOH and this model was taken to forward selection three as Fit3.

The third round of forward selection involved one genomic factor (17q22-17q24.1 CNN-LOH) as well as patient age at diagnosis and number of copy number abnormalities. Only the addition of patient age to Fit3 improved the model and this was taken to the fourth round of forward selection as Fit4.

Round four of forward selection determined whether addition of the number of copy number aberrations or 17q22-17q24.1 CNN-LOH would add to the Fit4 model. Neither model was significantly superior to the previous model of Fit4 which is considered the most appropriate model (Table 4.5, Figure 4.18A-B).

Risk Factor	**Time to progression (TTP)**	
	Hazard ratio (95% CI)	*p* value
3q29 gain	4.8016 (1.72-13.37)	0.003
17p13.3-17p11.2 CNN-LOH	4.3298 (1.51-12.40)	0.006
Age at Diagnosis	0.8836 (0.78-1.00)	0.049
17q11.2-17q21.32 CNN-LOH	4.5577 (1.42-14.64)	0.011

Table 4.5 - Multivariate survival analysis by Forward Selection Method revealed Fit4 to be the best predictive model of poor outcome in the cohort.

Within this model all four factors were independent significant prognostic markers. The three genomic markers of 3q29 gain, 17p CNN-LOH and 17q CNN-LOH all warrant further investigation to determine what gene within the region is driving disease resistance.

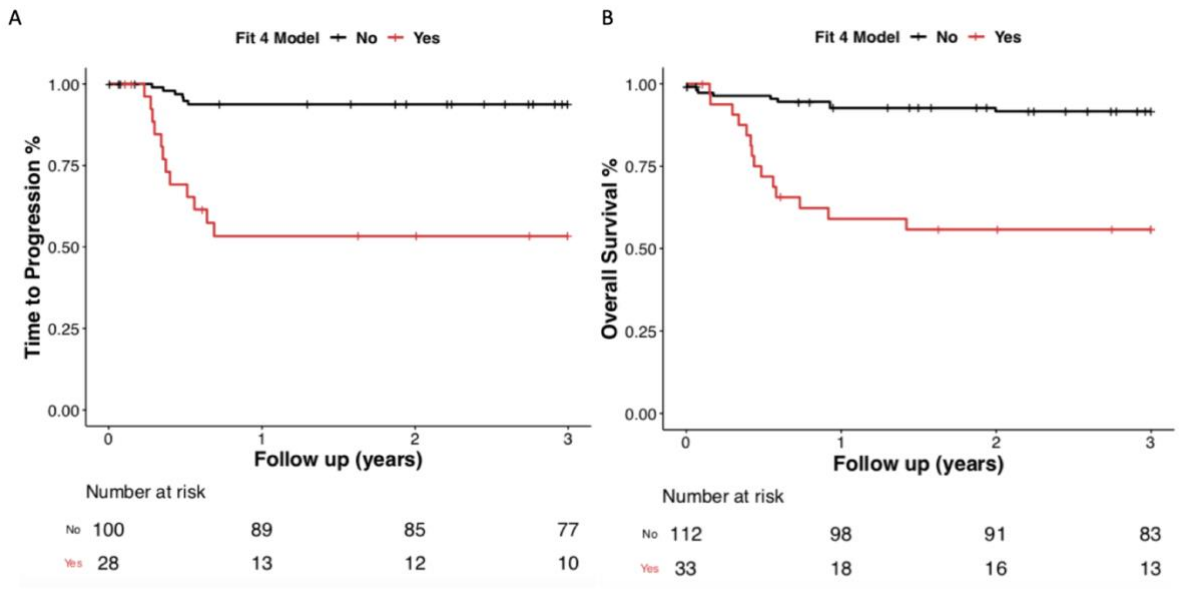


Figure 4.18 - Kaplan Meier plots demonstrating the Fit 4 multivariate risk model by (A) Time to Progression (p<0.001) and (B) Overall survival (p<0.001) analyses.

Applying the model to the cohort, 39 cases fit the criteria for the high-risk group of the model in having 3q29 gain, 17p13.3-17p11.2 CNN-LOH or 17q11.2-17q21.32 CNN-LOH. While a younger patient age was determined to be a significant factor in disease progression, with 14/18 cases who experienced relapsed/refractory disease being aged 10 or under, this cut-off led to 71.0% (115/162) cases being in high risk arm of the model, of which only 15.7% (18/115) had relapsed/refractory disease. This did however cover all cases that underwent disease progression. Removing patient age from the model reduces the number of patients in the high-risk group to 39 (24.1% of the cohort), including 12 cases with disease progression (Figure 4.19). Further validation in a larger cohort of patients is required to better tune the model so it is sensitive enough to pick up all cases with disease progression but specific enough to exclude the majority of cases that would be cured by frontline therapy. Investigating mutations within the regions of interest in the model, particularly regions of CNN-LOH on chromosome 17, may allow for improvement of the model. This is further investigated in Chapter 5.

4.3.11 Investigation of previously identified genomic prognostic markers

In addition to regions highlighted in our cohort by GISTIC2.0 analysis, prognostic markers reported by previous studies were investigated to determine whether their prognostic relevance was validated in our patients. Three genomic abnormalities were investigated: 7q21-7q22 gain, 13q34 deletion and 13q31.3 gain. Remarkably, no previously identified prognostic markers were associated with patient outcome or disease progression in our cohort (Appendix B). 7q22.1 gain reported in both Poirel *et al.* (2009) and Nelson *et al.* (2010) were present in 37/162 patients but were not enriched for in the disease progression group. Equally, while 13q34 loss was seen in 19/162 (11.8%) patients, the majority of these were associated with 13q complex aberrations which were not associated with survival. Equally, 13q31.3 gain involving the *MIR17HG* miRNA cluster was recurrent in the cohort but not associated with poor outcome or disease progression.

Explaining the disparity in prognostic markers between our cohort and those previously published is difficult. All three markers investigated were identified in patients uniformly treated with FAB/LMB96-based protocols, so it is unlikely that the disparity arose through different selection pressures through treatment. The abnormalities were also initially identified as prognostic in paediatric cohort, meaning that age is unlikely to play a role. 7q21-7q22 gain and 13q34 deletion were identified through traditional karyotyping, while 13q31.3 was identified by copy number array, as used in this study. It is possible that the small cohort size in the Schiffman *et al.* study led to *MIR17HG* gain being called a prognostic marker inappropriately. Our findings in this study validate the presence of 13q31.3 copy number gain in paediatric B-NHL, but dismiss the possibility of being a prognostic marker. Differences in the limitations of traditional karyotyping and copy number arrays may go some way to explaining the differences between studies. 7q21-7q22 gain was usually the result of a gain of the whole arm, but several events were smaller and might not have been picked up by karyotyping. These differences highlight how important it is to secure a clinical trial to validate putative prognostic markers in a large cohort of uniformly treated patients with complete clinical data.

4.3.12 Whole-exome sequencing analysis of 3q29 and 17q

39 cases with whole-exome sequencing data from tumour and matched normal DNA were processed and aligned as described in the methods section. Variants were called from the aligned BAM files using MuTect2 as part of GATK3.8 as described in Chapter 2 (Cibulskis *et al.*, 2013). The output files from variant effect predictor (VEP) were merged for 39 paired samples and variants were filtered for mutations on 17q and 3q29.

On chromosome 17q a total of 780 somatic variants were identified in 25/39 patients. After filtering out non-coding and synonymous variants, a total of 283 mutations in 13 patients were observed. 188 unique genes were mutated within the cohort and 57 were mutated more than once. The most frequently mutated gene on 17q was *RNF213*, with 12 variants called in seven unique cases. The *RNF213* gene on cytoband 17q25.3 codes for an E3 ubiquitin-protein ligase and has been implicated in cancer previously. The gene is a known fusion partner of *ALK* in anaplastic large cell lymphoma (van der Krogt *et al.*, 2017) and plays a role in hypoxia sensitivity in *HER2*+ breast cancer (Banh *et al.*, 2016). *RNF213* has not been implicated in B-NHL previously, and is not located within one of the two regions of 17q identified as having recurrent CNN-LOH by GISTIC2.0.

One BLL-11q case, 12/854 harboured 5 unique missense variants in *RNF213* of which one was present in the COSMIC database (Tate *et al.*, 2019). Despite this, all variants were predicted to be benign by PolyPhen, a tool which evaluates the effect of amino acid changes caused by somatic variants on protein function from a structural and evolutionary perspective (Adzhubei *et al.*, 2013) (Appendix C). BL patient 14/325 had two somatic missense mutations, of which one was flagged as probably damaging by PolyPhen, suggesting that the variant had a significant effect on *RNF213* protein function.

GNA13 at cytoband 17q24.1 was recurrently mutated in the cohort, as would be expected given the established role for mutations in the gene in B-NHL (Love *et al.*, 2012; Schmitz *et al.*, 2012; O'Hayre *et al.*, 2016). 9/39 (23.1%) patients were shown to harbour mutations of the *GNA13* gene including 8 single nucleotide variants and one 10bp deletion leading to a frame shift. No patient harboured more than one variant in the *GNA13* gene. Of eight single nucleotide variants identified three were nonsense mutations introducing a stop codon. Of the five missense mutations, four

were flagged as “probably deleterious” by PolyPhen while one was considered benign by PolyPhen. SIFT, an alternative functional predictor of non-synonymous mutations, considered all five missense mutations to be deleterious. This is consistent with the literature, where *GNA13* inactivating mutations have repeatedly been reported (Love *et al.*, 2012; O’Hayre *et al.*, 2016). Of two patients with paired exome sequencing data and 17q CNN-LOH of the region spanning *GNA13*, one had a mutation. Investigation of the association between 17q CNN-LOH and *GNA13* mutation was performed in an additional cohort of 34 cases without exome sequencing data from matched normal DNA. Seven non-synonymous *GNA13* mutations were observed in the cohort in 7/34 (20.6%) cases, including three nonsense mutations. It was not possible to eliminate the possibility of germline mutations being identified in this cohort, however only one of seven mutations observed was present in dbSNP, a database of previously reported variants (Sherry *et al.*, 2001). When considering all identified *GNA13* mutations and integrating the data with 17q CNN-LOH events it was determined that of six cases with 17q CNN-LOH spanning *GNA13*, four had mutations within the gene. This suggests that *GNA13* may be the driver gene within 17q CNN-LOH conferring poor outcome in this B-NHL cohort. Cox Proportional Hazard and Kaplan-Meier analysis was performed to determine whether *GNA13* mutations were associated with poor outcome (Figure 4.19). Cox Proportional Hazard method confirmed that *GNA13* mutations were not a prognostic in this cohort by TTP or OS (TTP: HR 1.4 95% CI 0.4-5.3, p=0.59, OS: HR 1.4 95% CI 0.3-7.0, p=0.68, Cox proportional hazard method).

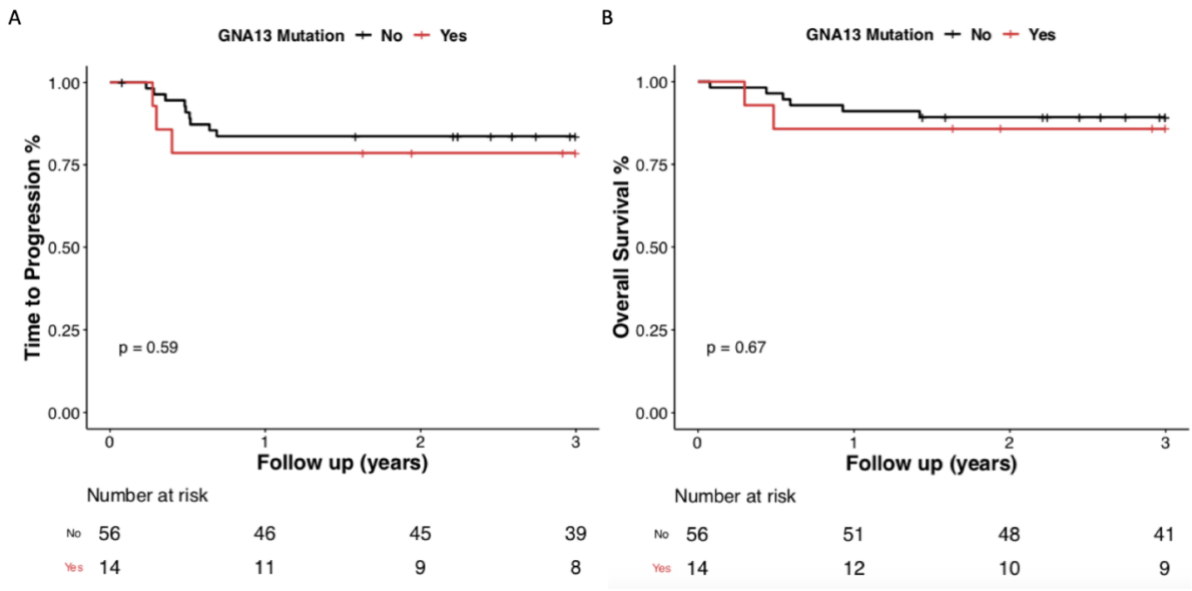


Figure 4.19 - Kaplan Meier plots demonstrating the risk associated with GNA13 mutation by (A) Time to Progression (p=0.587) and (B) Overall survival (p=0.674) analyses.

Conversely to *GNA13*, 55 coding mutations in *RNF213* were observed in the unpaired cohort of 34 cases, but only two were considered to be real as one was not present in dbSNP, while the other was present in dbSNP but also reported in COSMIC. No other genes were sufficiently recurrently mutated in the matched tumour-normal cohort for survival analysis on 17q. Expansion of the cohort with new samples is required for further investigation of genetic mutations associated with the prognostic marker 17q CNN-LOH.

3q29 gain was shown to be a significant prognostic marker by TTP and OS analysis. Present in 17/162 (10.5%) cases, 3q29 represents an opportunity for the identification of drug targets as well as being a prognostic marker with potential clinical relevance. Analysing exome sequencing data from 39 cases with matched tumour and normal DNA, 67 coding mutations were identified in 24 patients. After filtering out non-coding and silent mutations the number of mutations was 24 in 17 cases. A total of 15 genes were mutated in at least once case in this cohort, and four genes were mutated more than once: *ATP13A4* (n=2), *MUC4* (n=2), *PAK2* (n=4) and *OPA1* (n=5). Unlike 17q CNN-LOH where the aberration would be expected to co-occur with mutations, as 3q29 was amplified it is possible that mutations may drive the expression of a driver gene which would otherwise be overexpressed by a gene dosage effect as a result of additional copies of the gene. Out of the 39 cases with paired exome data, four harboured 3q29 gain. Of these, 2/4 had mutations of one of

the four recurrently mutated genes. Both cases underwent disease progression, including one BL with an *OPA1* mutation, and one DLBCL with a *PAK2* mutation (Appendix D).

Expanding this analysis to the remainder of the samples with exome sequencing data, *MUC4* was recurrently mutated in these samples, with 34/34 (100.0%) of cases having at least one mutation. Recurrent variants at chr3:195515594 (rs13095016) and chr3:195517258 (rs1106502) were seen in 32 and 33 patients respectively, and so they were removed from analysis as they likely represented germline mutations. *PAK2* and *OPA1* mutations were not observed in this cohort. One mutation in *ATP13A4* was observed in BL patient 11/634. It is noteworthy that the two most recurrently mutated genes in the paired samples were not observed in the unpaired cohort and so coding mutations in these genes of interest were not frequent enough to investigate via survival analysis. It is possible that non-coding mutations on 3q29 drive progressive disease in paediatric B-NHL, therefore whole-genome sequencing should be performed to screen for the presence of these variants and better understand the role of 3q29 in B-NHL.

4.4 Discussion

Despite increasingly high cure rates in paediatric B-NHL using intensive multi-agent chemotherapy, the outcome for patients who are not cured with frontline treatment is dire. Relapse and refractory disease are extremely difficult to predict at diagnosis, and salvage therapy for those who have undergone disease progression is largely ineffective. High resolution copy number analysis of large, clinically annotated datasets is required to identify significant markers of poor outcome in this particularly small, high-risk subgroup of patients.

The results from this study represent the largest copy number investigation of paediatric B-NHL to date, taking into account clinical data and patient outcome. Previous studies have identified recurrent abnormalities which may play a role in disease progression and relapse but none of these have been successfully translated into the clinic (Poirel *et al.*, 2009; Nelson *et al.*, 2010; Schiffman *et al.*, 2011). This study successfully identified three significant markers of disease progression that can be detected at diagnosis and further research into these aberrations may shed light on important drug targets.

Copy number gain of 3q has been reported in B-NHL before but studies have previously assumed that *BCL6* gain at 3q27.3 is the driver gene within this aberration (Scholtysik *et al.*, 2010). However, *BCL6* was gained in only 13/17 (76.5%) patients with gain at 3q29. Instead, it believed that other previously characterised oncogenes such as *PAK2* within the 3q29 locus drive disease progression in paediatric B-NHL patients with 3q29 gain. Mutations were recurrently identified in *PAK2*, *MUC4* and *OPA1* but at low frequencies. There was no association between 3q29 gain and mutations in genes within the region, suggesting either that amplification of a candidate gene drives disease progression alone, or that variants involved with progressive disease could not be picked up by whole exome sequencing alone.

17q CNN-LOH has not been thoroughly investigated in paediatric B-NHL previously despite the relatively high frequency identified in this study. Unlike in 17p CNN-LOH analysis, 17q does not immediately suggest the role of a particular tumour suppressor or oncogene. The most likely candidate from consideration of the literature was *GNA13*, which is found within the most frequently observed region of 17q CNN-LOH and the gene has a known role in both BL and GCB-DLBCL

pathogenesis. *GNA13* mutations are largely nonsense mutations which correlates with the findings of exome sequencing analysis in this study. Nonsense mutations and frameshift mutations cause abrogation of the gene, resulting in no detectable expression of the Gα13 G protein coupled receptor (GPCR) subunit it encodes. Aberrations involving the *GNA13* are present in between 10-25% of BL and GCB-like DLBCL cases (Healy *et al.*, 2016; O'Hayre *et al.*, 2016) but mutations were not associated with inferior TTP or OS. Whole-genome sequencing analysis should be performed to investigate the prognostic value of *GNA13* mutations and determine whether non-coding mutations are recurrent on 17q that confer an inferior prognosis.

The significant association between 17p CNN-LOH and disease progression is a strong indicator for a pathogenic role for *TP53* mutations and aberrations in paediatric B-NHL. This finding warrants further investigation of *TP53* by sequencing to determine whether mutations alone could be used as a clinical test for patient risk, which is detailed in the next chapter. *TP53* is an important tumour suppressor gene that is aberrant in about half of all cancer diagnoses and has been shown to be associated with poor outcome across multiple cancer subtypes (Robles *et al.*, 2016). Despite this, aberrant *TP53* status does not represent a clinically viable drug target. Establishing *TP53* status at diagnosis may be useful in stratifying patients into clinically relevant subgroups in the future. The role of *TP53* in paediatric B-NHL is characterised thoroughly in Chapter 5.

Chapter 5: Analysis of TP53 Abnormalities in Paediatric B-NHL.

5.1 Introduction

Copy number analysis of 162 paediatric B-NHL cases highlighted that abnormalities involving chromosome arm 17p, namely deletion of the arm and copy number neutral loss of heterozygosity (CNN-LOH), are recurrent in the disease, as demonstrated in Chapter 4. CNN-LOH of 17p has been reported previously (Sobol *et al.*, 2002; Scholtysik *et al.*, 2010; Schiffman *et al.*, 2011), as have deletions of 17p involving the *TP53* gene (Poirel *et al.*, 2009; Scholtysik *et al.*, 2010; Schiffman *et al.*, 2011). Schiffman *et al*. reported an association between amplification at 13q31.3 involving the *MIR17HG* miRNA cluster and 17p aberrations, with 3/3 cases harbouring either 17p deletion or CNN-LOH. Strikingly, 17p CNN-LOH was associated with significantly inferior time to progression (TTP) and overall survival (OS). Given that 17p harboured the *TP53* gene locus it was postulated that aberrations affecting the *TP53* gene may drive resistant disease in paediatric B-NHL.

The *TP53* gene has been shown to act as both a tumour suppressor gene and an oncogene (Soussi and Wiman, 2015). The gene is deleted and/or mutated frequently in most cancer subtypes and in about half of all cancer cases (Hollstein *et al.*, 1991), lending to the consideration that the gene acts as a tumour suppressor gene in some contexts. However, several mutant p53 proteins have been shown to be active as a result of gain-of-function mutations in the *TP53* gene (Yamamoto and Iwakuma, 2019), and mutations in the gene are most frequently missense and rarely frame shift or nonsense mutations, which is atypical of a tumour suppressor gene (Baugh *et al.*, 2018).

The p53 protein plays multiple roles in inhibiting inappropriate proliferation and initiating apoptosis in response to cell stress and extrinsic signalling. *TP53* gene mutations are predominantly missense mutations, in contrast to the vast majority of tumour suppressor mutations which are typically nonsense frameshift mutations (Petitjean *et al.*, 2007). Missense mutations in *TP53* are recurrently identified in multiple loci, another distinction from mutational patterns seen in many other tumour suppressors where a small number of mutation hotspots are observed. Analysis of the comprehensive *TP53* database maintained by the International Agency for Research on Cancer (IARC) (Bouaoun *et al.*, 2016) highlights that the most common

missense variants identified in cancer have a dominant-negative effect (DNE) in which the mutated transcript acts antagonistically against the wild type resulting in a loss of function. The vast majority of *TP53* mutations sit within the DNA-binding domain between codons 101 and 292 (Petitjean *et al.*, 2007). Mutations within this region are overwhelmingly missense mutations while these only make up approximately 40% of variants outside this region with a higher proportion of nonsense and frameshift variants. There are 22 CpG dinucleotide sites in the DNA binding domain of *TP53* and mutations of three codons make up 60% of missense variants reported in *TP53* CpG sites: R175, R248 and R273. In addition, p53 function can be abrogated indirectly, independent of *TP53* gene mutation. Amplification of *MDM2* and deletion of *CDKN2A* are both frequently reported in cancer and both are known to inhibit the p53 pathway (Alhejaily *et al.*, 2014).

Somatic *TP53* mutations in cancer has been studied for a long time (Hollstein *et al.*, 1991), however in recent years there has been a growing interest in germline *TP53* alterations. Li-Fraumeni Syndrome (LFS) was formally described in 1988 and characterised by a familial autosomal-dominant inherited pattern of various cancers in young patients (Li *et al.*, 1988). The increased susceptibility to cancer was later shown to be caused by germline mutations in *TP53* in five identified affected families (Malkin *et al.*, 1990). LFS differs from other hereditary cancer syndromes in the wide range of cancer diagnoses observed, primarily brain, bone and breast tumours. Patients with LFS are also more likely to acquire radiotherapy-associated secondary malignancies, highlighting a clinical consideration for treatment and monitoring of affected patients. Additional families with Li-Fraumeni-like syndromes (LFL) have been identified with germline *TP53* mutations but different clinical characteristics to classical LFS (Olivier *et al.*, 2003). In characterising these syndromes an association between *TP53* DNA binding domain and cancer diagnosis was observed, with DNA minor groove-binding mutations being associated more strongly with brain tumours (Olivier *et al.*, 2003). The study of the LFS/LFL using the IARC *TP53* database did not find specific germline mutations that were strongly associated with predisposition to lymphoma.

Leventaki *et al.* have previously reported that many paediatric BL patients harbour no aberrations in the *TP53* other than increased expression of *MDM4*. The investigators used *CDKN1A* immunohistochemistry (IHC) to measure impairment of the *TP53*

pathway and showed that *CDKN1A* as expressed in no cases. The authors hypothesised that while *TP53* mutation and *MDM2* overexpression and amplification acted to impair *TP53* function in a number of cases, overexpression of the *MDM4* gene achieved the same effect in the remainder of the cohort. *MDM4* has been shown to downregulate *CDKN1A* expression by abrogating the transactivation function of *TP53* (Marine *et al.*, 2006). *TP53* mutations were identified in 18% of cases in the study which was consistent with the 20% of cases identified in a sequencing study of the same year (Love *et al.*, 2012; Schmitz *et al.*, 2012).

TP53 status is assessed and used clinically in several cancers to guide treatment decisions, most notably in chronic lymphocytic leukaemia (CLL) (Campo *et al.*, 2018). In CLL both *TP53* mutation and deletion of 17p involving the *TP53* gene are independent prognostic markers of chemotherapy resistance. *TP53* aberrations were only observed in 10% of treatment-naïve CLL patients at diagnosis, but were more common in CLL samples that were refractory to fludarabine (Zenz *et al.*, 2009). It was determined that *TP53* aberrant CLL cases are effectively treated by BCL2 inhibitor Venetoclax (Stilgenbauer *et al.*, 2016), PI3K inhibitor Idelalisib (Brown *et al.*, 2014) and BTK inhibitor ibrutinib, leading to a marked improvement in the survival of these patients. The relevance of these inhibitors outside in *TP53* aberrant cancer outside of CLL has not been characterised, however. *TP53* status has been shown to be a prognostic factor in multiple other cancer subtypes, highlighting an opportunity to improve outcome in a large number of patients. *TP53* status stratification in CLL is a recent success story in the field of personalised therapy for cancer, following in the wake of initiatives such as screening for somatic mutations in *EGFR* and *ALK* fusions in lung cancer and *BRAF* mutations in melanoma patients (Kelleher *et al.*, 2012). The major distinction between these examples and *TP53* stratification is that *TP53* mutation is not targetable in itself, but instead defines a subgroup of patients who benefit from alternative chemotherapy.

5.2 Chapter Aims

The specific aims of the work detailed in this chapter were as follows:

- Investigate and characterise copy number aberrations and mutations affecting the *TP53* gene in paediatric B-NHL.
- Investigate the association between *TP53* abnormalities on disease progression individually and in an integrated analysis.
- Identify recurrent genomic abnormalities associated with *TP53* status and test for prognostic significance.

5.3 Results

5.3.1 Patient and Clinical Demographics

The *TP53* status of the diagnostic tumour samples from 95 UK paediatric B-NHL patients with copy number and mutation data for the *TP53* gene was analysed (Appendix E). Due to a lack of *TP53* mutation data cases from the literature that were analysed in Chapter 4 were excluded from analysis in this section of the project. Clinical follow-up data were available for 88/95 (92.6%) cases. *IG-MYC* translocations were detected in 68 of 95 (71.6%) cases, including 60 BL, 4 DLBCL and 4 B-NHL, NOS. Six cases failed FISH or had no available material and 21 had no detectable *MYC* translocation. In total, 88 of 95 (92.6%) patients had complete clinical follow-up data; 77/88 (87.5%) were treated on FAB/LMB 96 protocols (trial or interim guidelines), one received no treatment, four were treated with alternative NHL protocols and treatment data were unavailable for 6 cases.

5.3.2 Copy number aberrations involving the TP53 locus

Following on from copy number analysis performed on 162 patients in Chapter 4, abnormalities involving 17p were interrogated in this cohort of 95 patients. 17p deletion involving *TP53* was identified in 15/95 (15.8%) cases with median region of loss of 20.6 Mb (range 17.9-22.2 Mb), resulting in deletion of >80% of 17p in all cases. Median log ratio of deleted cases was -0.40 (range -0.14--0.79). All cases with 17p loss had one copy deleted and there were no homozygous deletions of 17p. Additionally, 8/95 cases (8.4%) showed CNN-LOH with a median region of CNN-LOH

of 19.5 Mb (range 12.9-22.3 Mb) covering >80% of 17p (Figure 5.1). All 17p deletions and CNN-LOH involved the *TP53* locus.

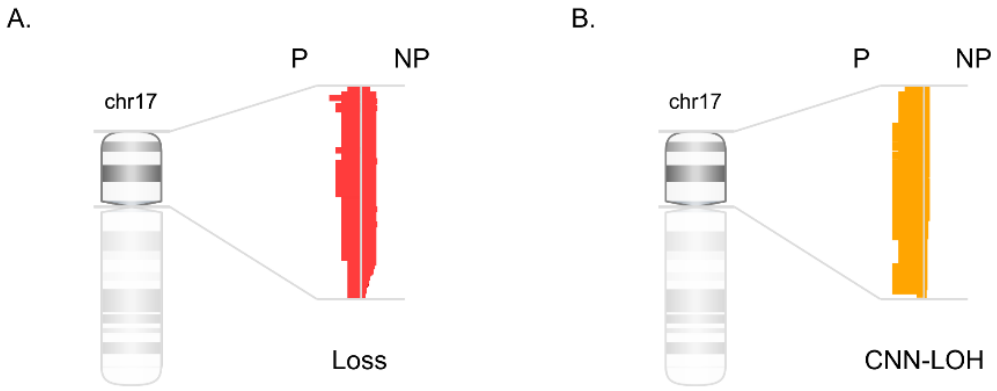


Figure 5.1 – (A) Deletion and (B) Copy Number Neutral Loss of Heterozygosity of chromosome arm 17p in 95 paediatric B-NHL diagnostic samples. P – Progression, NP – No Progression.

17p CNN-LOH was shown to be a highly recurrent abnormality in this cohort, strongly suggesting that mutations in a gene on this chromosome arm plays a role in driving therapy resistance in paediatric B-NHL.

5.3.3 Genomic Analysis of the TP53 locus

The *TP53* mutation status of all 95 cases was determined by whole-exome sequencing (n= 90) or Sanger sequencing (n=5). Variants were called by MuTect and Mutect2 variant callers and manually validated by in the Integrative Genomics Viewer software. For patients that did not have whole exome sequencing data but for whom there was patient DNA still available, Sanger sequencing of exons 5-8 was performed as described in Chapter 2.

In all, 56 variants were identified in the cohort of 95 paediatric B-NHL patients. Non-synonymous *TP53* mutations were found in 46/95 (48.4%) cases: 37 had a single non-synonymous somatic mutation, 8 had two mutations and one had a germline R248Q mutation inherited variant that was observed in the tumour and constitutional DNA which corroborated clinical data showing the patient had Li-Fraumeni Syndrome. One case harboured a single R213R mutation that was excluded for being a silent variant, while another harboured the same mutation as well as a missense E258D variant and was considered to have a monoallelic *TP53* abnormality as the silent mutation is assumed to have no functional consequence.

Published data from the UMD and IARC *TP53* Databases showed all but 2 of the 54 non-synonymous mutations to be functionally deleterious (Leroy *et al.*, 2014; Bouaoun *et al.*, 2016) (Appendix E). As seen in other cancers (Bouaoun *et al.*, 2016; Donehower *et al.*, 2019), the most frequently mutated residues were R175 (n=6), G245 (n=5), R248 (n=7) and R273 (n=4) (Figure 5.1). Of the 56 mutations identified in this cohort, 55 appeared to cluster into one of three regions within the DNA-binding domain: 1. V122-R158, 2. V173-V126, 3. S241-L289 (Figure 5.2).

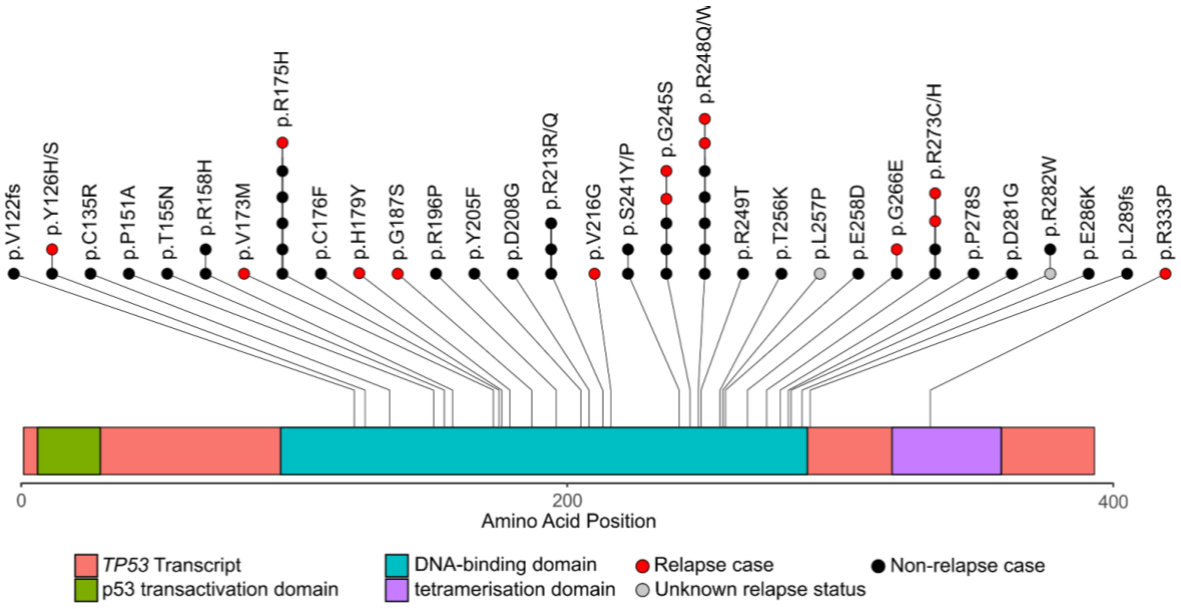


Figure 5.2 - TP53 Mutations identified in 46/95 paediatric B-NHL patients.

5.3.4 Integrated Analysis of TP53 Abnormalities in Paediatric B-NHL

Integrating the mutation and copy number array data showed that abnormalities involving the *TP53* gene were common at presentation with 52/95 (54.7%) cases harbouring at least one abnormality involving *TP53* (Figure 5.3). Cases with *TP53* aberration events that were believed to affect both alleles (termed “biallelic”) made up half of these, being present in 26/95 (27.4%) cases (Figure 5.4). Among 26 cases with biallelic abnormalities at diagnosis, 17 had mutation together with deletion or CNN-LOH, two had homozygous mutations without CNN-LOH and seven had two heterozygous mutations that were assumed to be compound heterozygous, affecting both alleles (Figure 5.3, Figure 5.4). Compound heterozygous mutations of *TP53* have been reported in the literature and are shown to be situated on alternate alleles (Havelange *et al.*, 2016b; Sande *et al.*, 2018). Of the 26 diagnostic samples with

monoallelic *TP53* abnormalities, 20 had a single somatic heterozygous mutation (median variant allele frequency (VAF) 39%, range 18-50%), four had a deletion and two had CNN-LOH (Figure 5.3). Two cases with CNN-LOH were identified without a detectable mutation of the *TP53* gene, it was assumed that an undetected subclonal *TP53* mutation was present at a low level in a small subclone but could not be considered to have biallelic abnormalities of *TP53*. This does not explain the clonal distribution of 17p CNN-LOH, however. It is possible that non-coding mutations may contribute to disease progression in a similar manner to missense mutations in coding regions. Non-coding regions of the *TP53* were not covered by the whole exome sequencing kits used in this study. Non-coding variants affecting the *TP53* gene have been reported in cancer previously (Gyorffy *et al.*, 2018).

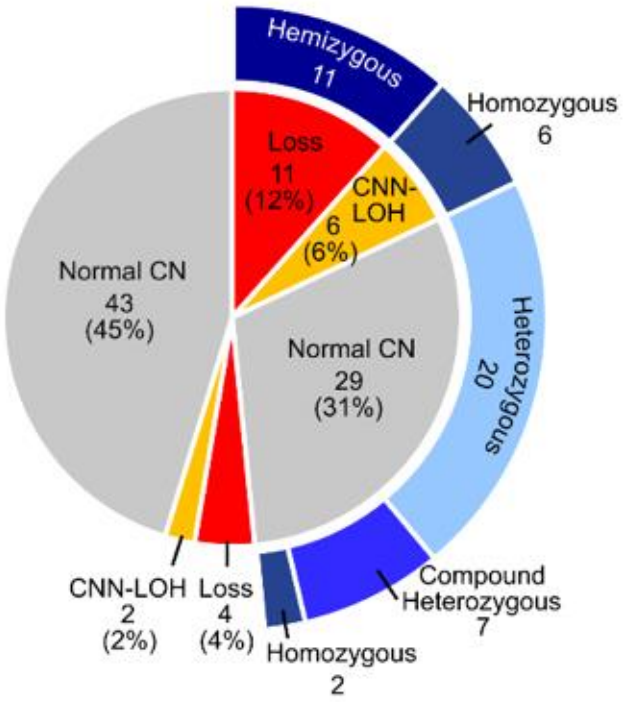


Figure 5.3 – A pie chart representation of co-occurrence of TP53 abnormalities in 95 paediatric B-NHL patients.

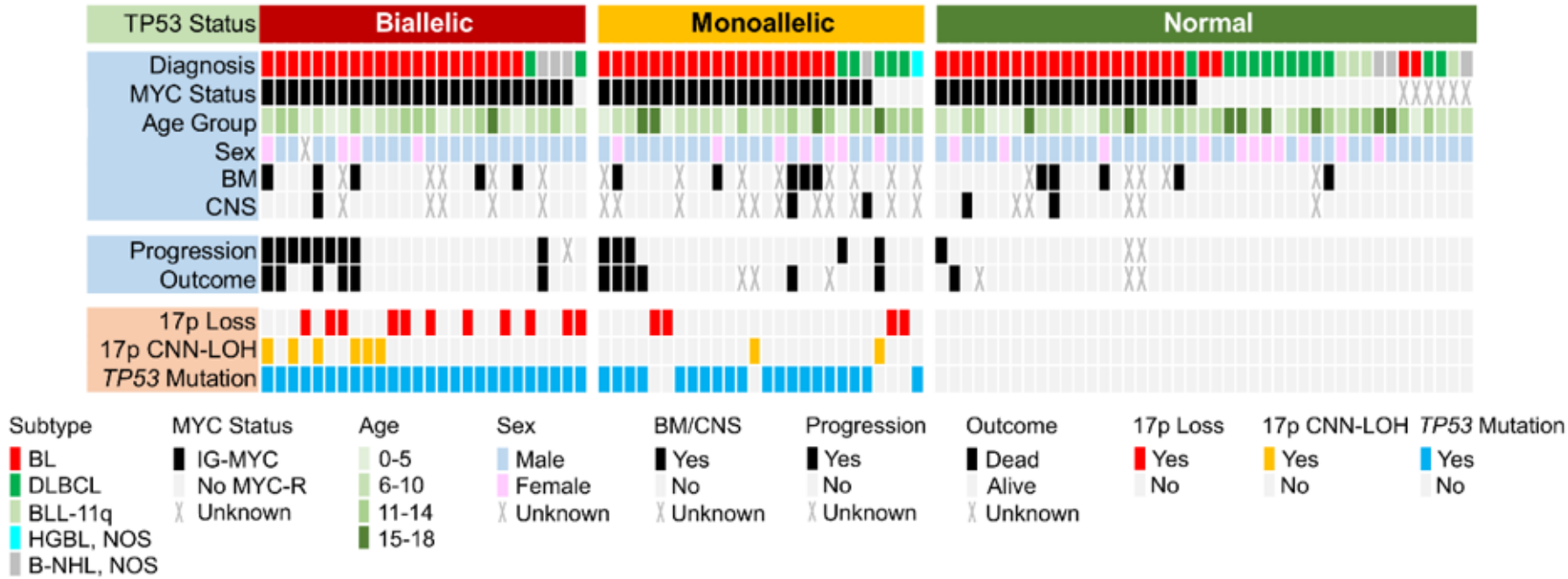


Figure 5.4 - Integrated analysis of TP53 abnormalities in paediatric B-NHL. An oncoplot showing TP53 status with clinical and molecular factors as described in the key. Data is plotted from left to right according to the presence of biallelic, monoallelic or no TP53 abnormality. Unavailable data is indicated with a cross.

Two cases had mutations that were less than 100bp apart and therefore could be covered in the same read (Figure 5.5). Both cases showed that the mutations were on alternating reads, with no reads containing both variants in either case. Based on this information, we could determine that these mutations affected both alleles, and therefore both cases had biallelic *TP53* abnormalities. The further six cases with compound heterozygous mutations were considered to have biallelic *TP53* aberrations based on this finding. Compound heterozygous cases are broken down in Table 5.1.

Case	**Variant 1**	**Variant 2**	**Distance between variants (bp, codons)**	**Confirmed compound heterozygous?**
14/325	R158H	R175H	52, 17	Yes
14/414	R175H	G245S	209, 70	No
15/156	S241Y	G245S	12, 4	Yes
15/314	R158H	P278S	359, 120	No
15/353	R248Q	R273C	74, 25	No
21/257	Y126H	G245S	357, 119	No
4/830*	L257P	R282W	74, 25	No
5/686	R248Q	T256K	24, 8	No

*Table 5.1 - Table showing the locations of variants in cases with compound heterozygous biallelic mutations in the TP53 gene. * Patient 4/830 also harboured a 17p deletion.*

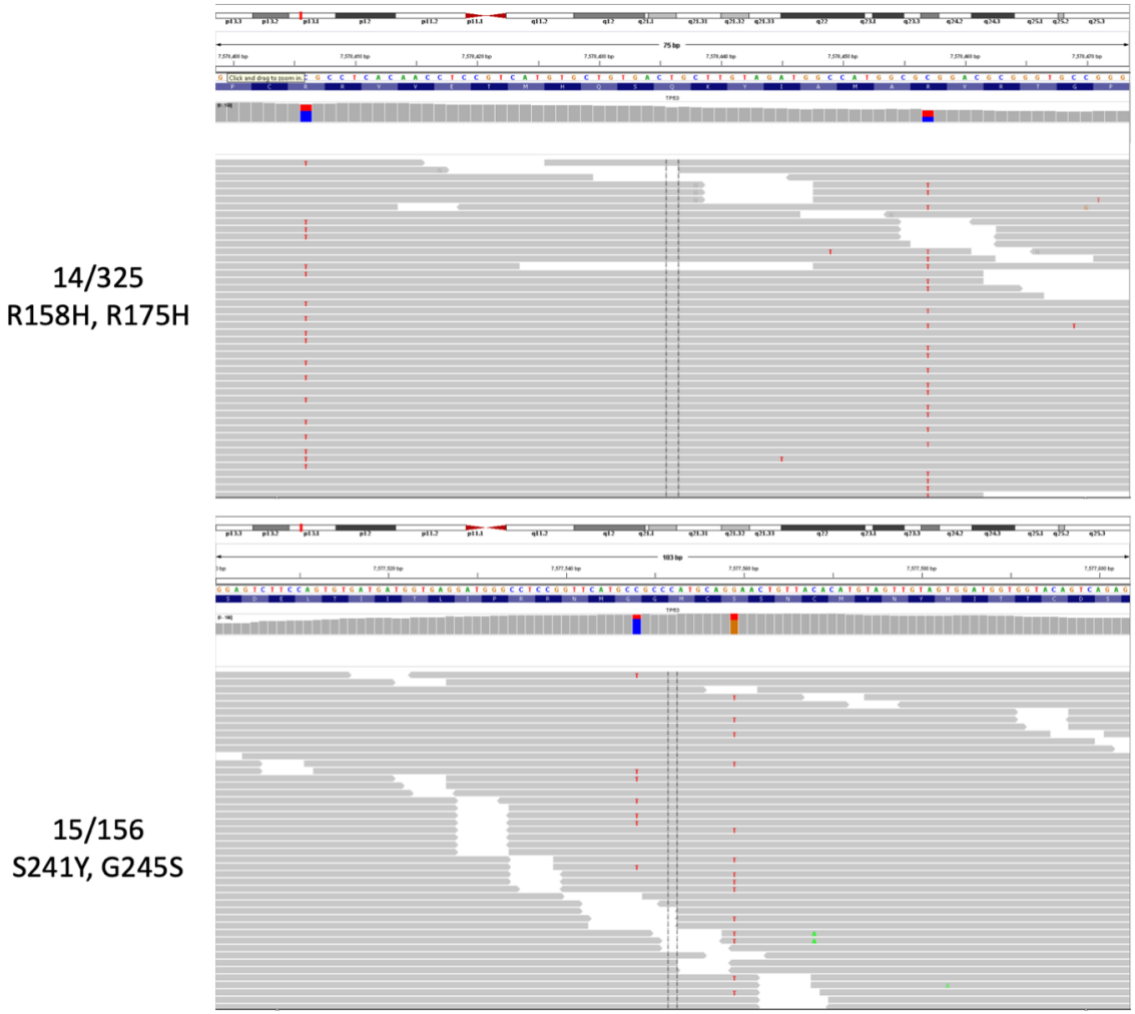


Figure 5.5 - Whole Exome Sequencing Data viewed in IGV showing two cases with compound heterozygous TP53 mutations. Grey horizontal bars represent individual whole exome sequencing reads covering the TP53 locus. Each case harboured two heterozygous mutations in the TP53 gene which could be covered by the same reads. In both cases there were no reads exhibiting both variants which strongly suggests they occurred on separate molecules.

Mutations in patient 14/325 both presented at high VAFs (54%, 36%) suggesting a high tumour content and that close to 100% of cells in the sample had both mutations. In contrast patient 15/156 presented with low VAFs (32%, 22%) which suggests that either the tumour content was low or that the compound heterozygous mutations were found in a subclone of the tumour sample. Patient 15/156 was designated in the pathology review as a B-NHL, NOS case, but was found to be 100% IG-MYC positive by FISH (100/100 cells, performed previously in lab).

5.3.5 Prognostic investigation of TP53 aberrations in paediatric B-NHL

Survival estimates at 3 years for overall survival (OS) and time to progression (TTP) were 85.2% (95% CI 78.1-93.0) and 82.4% (95% CI 74.6-90.9), respectively. For those patients with primary refractory or relapsed disease, the median time from initial diagnosis to disease progression was 4.8 months (range 2.8-8.2 months). Bone marrow (BM) and Central nervous system (CNS) involvement were associated with an inferior overall survival; BM involvement reached our significance threshold (HR 4.5 (CI 95% 1.3-16.0), p=0.018, Cox proportional hazard method) while CNS involvement almost achieved significance (HR 4.6 (CI 95% 1-22.0), p=0.057, Cox proportional hazard method).

Investigating the prognostic relevance of *TP53* in all 95 B-NHL cases by univariate cox modelling revealed that deletion of 17p did not confer an inferior TTP or OS independently (TTP HR 1.3 95% CI 0.4-4.8, p=0.650 and OS HR 0.4 95% CI 0.1-3.2, p=0.390, Cox proportional hazard method). However, 17p CNN-LOH was significantly associated with worse outcome (TTP HR 7, 95% CI 2.4-21.0, p<0.001 and OS HR 5.6, 95% CI 1.7-18.0, p=0.005, Cox proportional hazard method). 17p deletion involving *TP53* was observed twice as often as CNN-LOH of the same region (14/95, 15.9% vs 7/95, 8.0%, p=0.164, Fisher exact test) despite having no significant effect on TTP or OS. This suggested that mutations of the *TP53* gene may play a more important role in resistance to therapy than deletion.

Survival analysis by Cox Proportional Hazard model method demonstrated that *TP53* mutated cases had a significantly inferior TTP and OS than non-mutated cases (TTP HR 8.1, 95% CI 1.8-36.0, p=0.006 and OS HR 6.4, 95% CI 1.4-29, p=0.015, Cox proportional hazard method) (Table 5.2).

Variable	Time to progression (TTP)		Overall survival (OS)		
	HR (95% CI)	*p* value	HR (95% CI)	*p* value	No. cases with abnormality (%)
Univariate analysis					
CNS involvement	1.58 (0.2-12.4)	0.660	4.60 (0.95-22.0)	0.057	5/72 (6.9%)*
BM involvement	2.17 (0.7-7.2)	0.207	4.50 (1.3-16.0)	0.018	15/74 (20.3%)*
TP53 deletion	1.30 (0.4-4.8)	0.650	0.41 (0.1-3.2)	0.390	14 (15.9%)
TP53 CNN-LOH	7.00 (2.4-21.0)	0.00042	5.60 (1.7-18.0)	0.0045	7 (8.0%)
TP53 mutation	8.10 (1.8-36.0)	0.0058	6.40 (1.4-29.0)	0.015	43 (48.9%)
Any TP53 abnormality	14.00 (1.8-100.0)	0.012	11.00 (1.4-85.0)	0.021	48 (54.5%)
TP53 monoallelic abnormality	9.90 (1.2-85.0)	0.036	11.00 (1.4-94.0)	0.025	23 (26.1%)
TP53 biallelic abnormality	16.00 (2.1-130.0)	0.0081	10.00 (1.2-86.0)	0.031	25 (28.4%)

*Table 5.2 – Univariate survival analysis of TP53 abnormalities in paediatric B-NHL. * Data for BM and CNS involvement was not available for all cases with follow-up data.*

Strikingly, when these variables were combined, the presence of any *TP53* abnormality was strongly associated with a poor outcome (TTP HR 14 (95% CI 1.8-100.0), *p*=0.012, OS HR 11 (95% CI 1.4-85.0), *p*=0.021, Cox proportional hazard method) when compared to those without any abnormalities of *TP53* with higher hazard ratios by both survival metrics. The presence of a biallelic *TP53* abnormality at presentation was shown to be associated with a significantly inferior TTP than *TP53* normal cases (TTP HR 16.0, 95% CI 2.1-130.0, *p*=0.008, Cox proportional hazard method). Similarly, monoallelic aberrations of the *TP53* gene were also associated with a dire outcome (TTP HR 9.9, 95% CI 1.2-85.0, *p*=0.036, Cox proportional hazard method). This demonstrates that cases with *TP53* aberrations affecting both alleles represent the absolute worst genetic risk group in paediatric B-NHL identified to date. Kaplan Meier plots for TTP and OS demonstrate the poorer prognosis of biallelic *TP53* cases in comparison to monoallelic and normal cases (Figure 5.6).

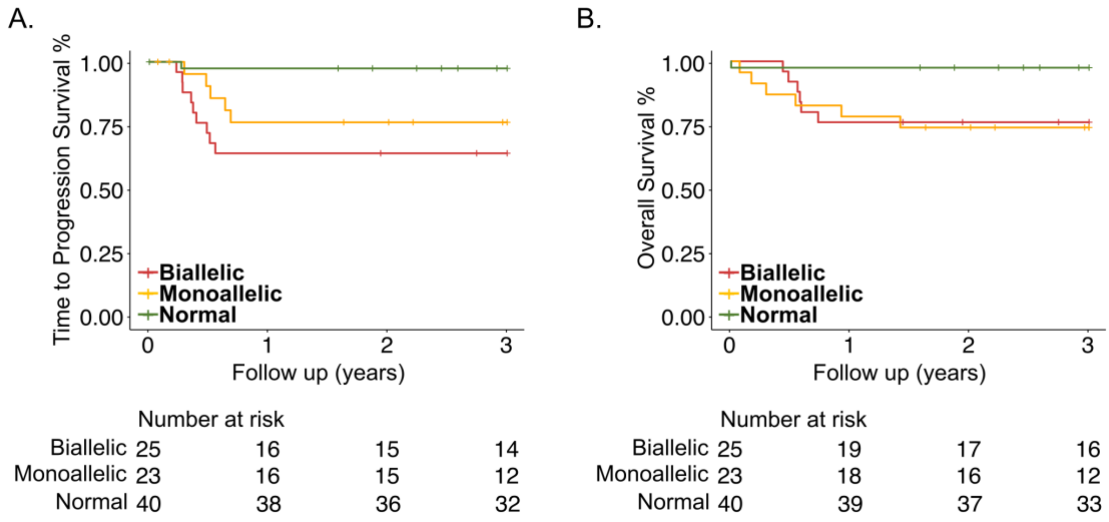


Figure 5.6 - Kaplan Meier plots showing the (A) Time to progression (Biallelic vs Normal p<0.001, Biallelic vs Monoallelic p=0.002, Monoallelic vs Normal p=0.010, Log rank test) and (B) Overall survival analysis of TP53 status in paediatric B-NHL. (Biallelic vs Normal p-0.007, Biallelic vs Monoallelic p=0.137, Monoallelic vs Normal p=0.005, Log rank test)

It is remarkable that only one patient who underwent disease progression resided within the *TP53* normal group (Figure 5.6A). This patient was analysed as a diagnostic and relapse pair in Chapter 7 and was shown to have biallelic *TP53* aberrations in the relapse biopsy, specifically deletion of 17p and a hemizygous R248W mutation. This means that all BL cases that underwent disease progression

harboured a *TP53* aberration. It is possible that the case with no detectable *TP53* abnormality at diagnosis harboured an undetected subclonal *TP53* mutation. Of those patients without any *TP53* abnormality at initial diagnosis had an OS of 97.5%, with one patient dying early during treatment (Figure 5.6B).

It was hypothesised that specific residues within the DNA-binding domain conferred different degrees of protection to the tumour against chemotherapy, and mutations within these residues may confer a poorer prognosis. Survival analysis did not show a significant association between disease progression or patient outcome for any cluster of mutation sites in particular.

5.3.6 Multivariate survival analysis investigating TP53 aberrations in paediatric B-NHL

Multivariate analysis was performed to assess the impact of *TP53* abnormalities alongside established high-risk clinical factors (BM and CNS involvement). Importantly, the presence of any *TP53* abnormality emerged as an independent prognostic marker in the paediatric B-NHL cohort (TTP HR 11.69 (95% CI 1.49-91.57), p=0.019, Cox proportional hazard method) (Table 5.3), in contrast to established clinical factors of BM and CNS involvement. Furthermore, biallelic *TP53* abnormalities were even more strongly associated with disease progression (TTP HR 14.43 (1.76-118.02), p=0.013, Cox proportional hazard method), than monoallelic *TP53* abnormalities (TTP HR 8.20 (0.85-79.83), p=0.068, Cox proportional hazard method).

Risk Factor	**Time to progression (TTP)**	
	Hazard ratio (95% CI)	p **value**
B-NHL (n=88 B-NHL cases with genomic and survival data)		
Multivariate analysis - Biallelic and monoallelic abnormalities		
TP53 monoallelic abnormality	8.20 (0.85-79.83)	0.068
TP53 biallelic abnormality	14.43 (1.76-118.02)	0.013
CNS involvement	1.37 (0.16-11.78)	0.771
BM involvement	1.47 (0.36-5.90)	0.590

Table 5.3 - Multivariate survival analysis of TP53 abnormalities in paediatric B-NHL.

5.3.7 TP53 abnormalities by B-NHL subgroup

Subsequent survival analyses were performed to investigate whether the prognostic relevance of *TP53* abnormalities were associated with a specific subtype of B-NHL. The cohort in this part of the study was primarily made up of BL cases, and univariate analysis confirmed that both biallelic abnormalities and any *TP53* mutations were associated with a significantly poorer time to progression (Biallelic HR 8.80 (95% CI 1.1-70), p=0.041, any TP53 mutation HR 8.30 (95% CI 1.1-64), p=0.043, Cox proportional hazard method) in this subgroup (Appendix F). The presence of any abnormalities involving *TP53* was not a significant prognostic marker in paediatric BL but bordered on significance (HR 7.00 (95% CI 0.91-54), p=0.062, Cox proportional hazard method). The reduction in significance is likely attributed to the reduced statistical power as a result of a smaller cohort size. Kaplan Meier analysis by Log Rank highlighted that there is a correlation between *TP53* abnormalities and poor outcome by overall survival and time to progression analysis (Figure 5.7). The *TP53* normal group still do remarkably well in the BL subset, and the one case that underwent disease progression that was *TP53* normal at original diagnosis acquired *TP53* biallelic abnormalities at the point of relapse (described further in Chapter 7).

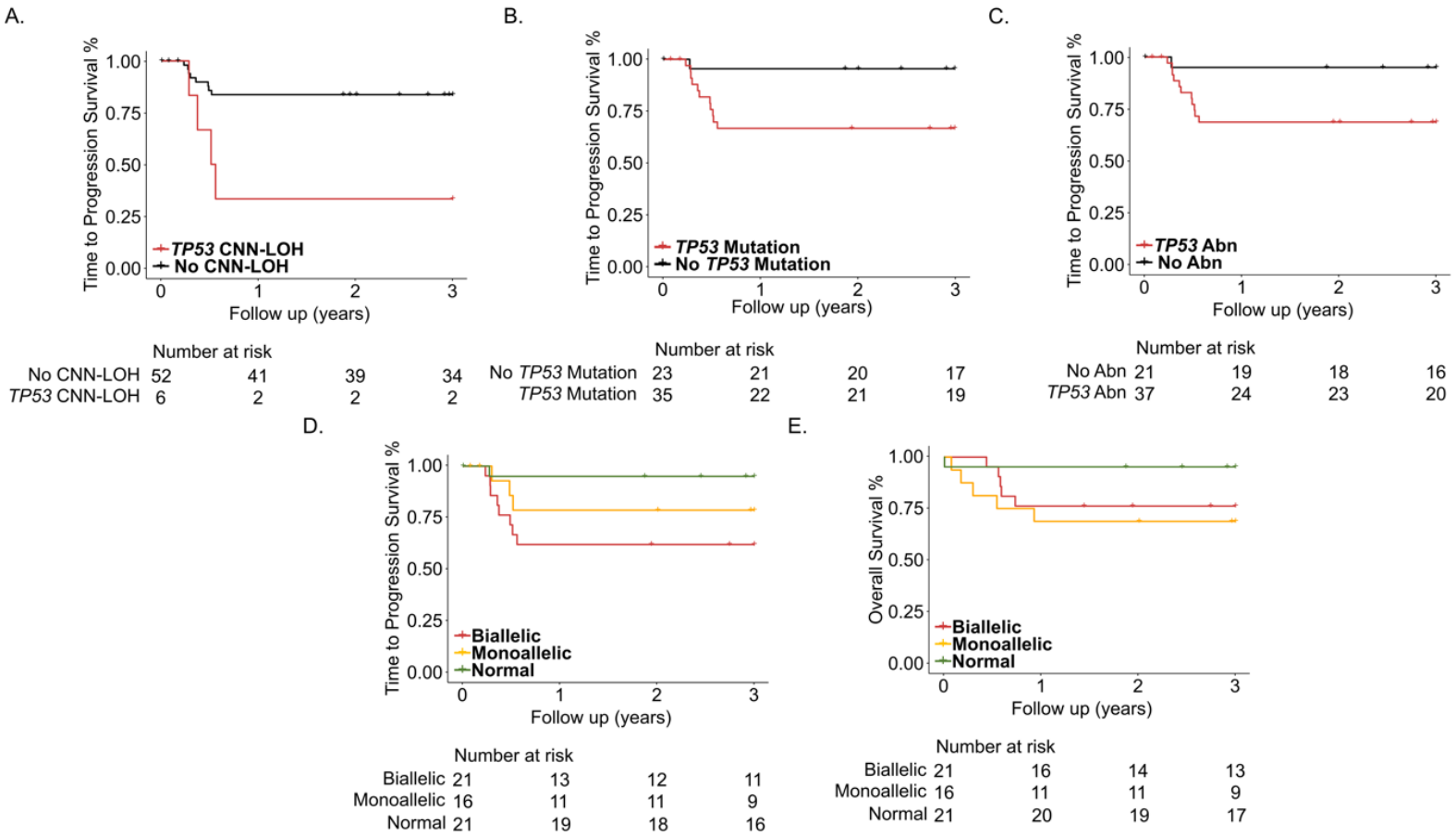


Figure 5.7 - Kaplan Meier plots showing cumulative survival of BL patients with TP53 abnormalities.(A) TTP Survival of CNN-LOH involving TP53 gene (p<0.001, Log rank test) (B) TTP Survival of TP53 Mutation (p=0.001, Log rank test) (C) TTP Survival of any TP53 abnormality (p=0.001, Log rank test) (D) TTP Survival of TP53 Status in BL (Biallelic vs Normal p<0.001, Biallelic vs Monoallelic p=0.002, Monoallelic vs Normal p=0.010, Log rank test) and (E) Overall Survival of TP53 Status in BL (Biallelic vs Normal p-0.007, Biallelic vs Monoallelic p=0.137, Monoallelic vs Normal p=0.005, Log rank test).

In an investigation of 19 DLBCL patients with outcome and copy number data there were four cases with *TP53* mutations, one of whom had Li-Fraumeni syndrome, with an inherited R248Q mutation. Out of the three cases with somatic *TP53* mutations, one was a heterozygous R248W mutation, one was a hemizygous R213Q mutation alongside a single deletion of 17p, and the other was a case that experienced disease progression and had a single heterozygous G187S mutation. The other DLBCL case with disease progression harboured 17p CNN-LOH that involved the *TP53* gene, but no mutations were identified in either the diagnostic or relapse samples via whole exome sequencing. There were too few cases to perform formal survival analysis, however of two patients who underwent disease progression, one had 17p CNN-LOH involving *TP53* albeit without a detectable mutation and the other harboured a G187S mutation.

5.3.8 p53 Pathway Defects and Abnormalities

A panel of genes commonly aberrated in cancer and associated functionally with p53 was identified using KEGG Ontology (Kanehisa *et al.*, 2019) and screened using both copy number arrays (in all 95 patients) and whole exome sequencing (in 39 patients with paired tumour and normal DNA sequencing) to determine whether additional defects in the p53 pathway were associated with or contributed to poor outcome in B-NHL (Table 5.4).

Gene	**Gain**	**Loss**	**CNN-LOH**	**Coding Mutation**
TP73	1/95 (1.1%)	3/95 (3.2%)	6/95 (6.3%)	0/39 (0.0%)
MDM4	15/95 (15.8%)	2/95 (2.1%)	2/95 (2.1%)	0/39 (0.0%)
MDM2	8/95 (8.4%)	1/95 (1.1%)	0/95 (0.0%)	0/39 (0.0%)
ATR	2/95 (2.1%)	3/95 (3.2%)	3/95 (3.2%)	3/39 (7.7%)
CDKN1A	1/95 (1.1%)	2/95 (2.1%)	9/95 (9.5%)	0/39 (0.0%)
CDKN2A	1/95 (1.1%)	11/95 (11.6%)	8/95 (8.4%)	0/39 (0.0%)
ATM	12/95 (12.6%)	0/95 (0.0%)	2/95 (2.1%)	2/39 (5.1%)
CHEK1	4/95 (4.2%)	7/95 (7.4%)	6/95 (6.3%)	0/39 (0.0%)

Table 5.4 - A table summarising TP53 pathway gene abnormalities in a cohort of paediatric B-NHL patients. Gain, loss and CNN-LOH were investigated in 95 patients with array data. Coding mutations were screened for in 39 patients with paired tumour and normal exome sequencing data.

Remarkably, mutations in these genes were found at very low incidence in a subset of 39 patients with paired tumour and normal exome sequencing data. Previous investigation into p53 pathway abnormalities in paediatric BL by Leventaki *et al.* (2012) suggested that abrogation of functional p53 protein occurs through multiple mechanisms, including *MDM2* amplification, but this was not observed. Somatic missense mutations in *ATM* were identified in two cases; one BLL-11q and one DLBCL case. Three cases harboured single missense mutations in the *ATR* gene, two of which were cases with mutated *ATM.* The additional *ATR* mutant case harboured an L1191F mutation and was a BL case that went on to relapse, 15/310. This patient had heterozygous G245S *TP53* mutation at diagnosis but no other *TP53* abnormality. No other coding mutations in p53 pathway genes were detected. *MDM4* copy number gains were recurrent, occurring in 15.8% of the cohort. The gene is located on 1q in a region of recurrent copy number gain in B-NHL as described in Chapter 4. Copy number gain of *MDM4* has been shown to correlate with high expression of both long and short forms of the MDM4 protein (Leventaki *et al.*, 2012).

5.3.9 B-NHL samples with TP53 abnormalities are associated with chromosomal complexity

Investigation of the genomic landscape of cases with biallelic *TP53* aberrations revealed that this subset of patients had more complex genomes than cases who were *TP53* normal and harboured complex chromosomal abnormalities (Table 5.5). Statistical analysis by one-way t-test confirmed that this association was significant (p=0.007, student t-test).

***TP53* Status**	**Mean % Genome Altered**	**Significance (p)**	**Mean CN Segments over 25kb**	**Significance (p)**	**Mean Complex Abnormalities per sample**	**Significance (p)**
TP53 Abnormal	6.44	0.3776	48.13	0.08502	1.038462	0.003633
Biallelic	6.04	0.4422	39.62	0.2328	1.576923	0.006518
Monoallelic	6.85	0.5304	56.65	0.162	0.5	0.1446
TP53 Normal	4.57	-	29.49	-	0.2790698	-

Table 5.5 - Investigation of the association between genomic complexity and TP53 status in paediatric B-NHL. P values were calculated using an unpaired t-test.

For an abnormality to be considered a complex chromosomal abnormality based on copy number array data it needed to exhibit fluctuation between two or more copy number states with two or more individual altered segments on one arm. The segments had to be at least 250kb in size to rule out common germline CNVs and technical noise. This includes abnormalities resembling the characteristic 11q abnormality seen in BL cases and MYC-negative BLL-11q cases, as well as instances of chromothripsis and complex stepwise gain similar to those seen in iAMP21 abnormalities seen in B-ALL cases (Rand *et al.*, 2011). iAMP21 (intrachromosomal amplification of chromosome 21) is a complex genomic abnormality that defines a subtype of paediatric B cell precursor acute lymphoblastic lymphoma (B-ALL). The mechanism driving this aberration has been characterised through extensive FISH analysis as a breakage-fusion-bridge mechanism followed in many cases by chromothripsis (Robinson *et al.*, 2007).

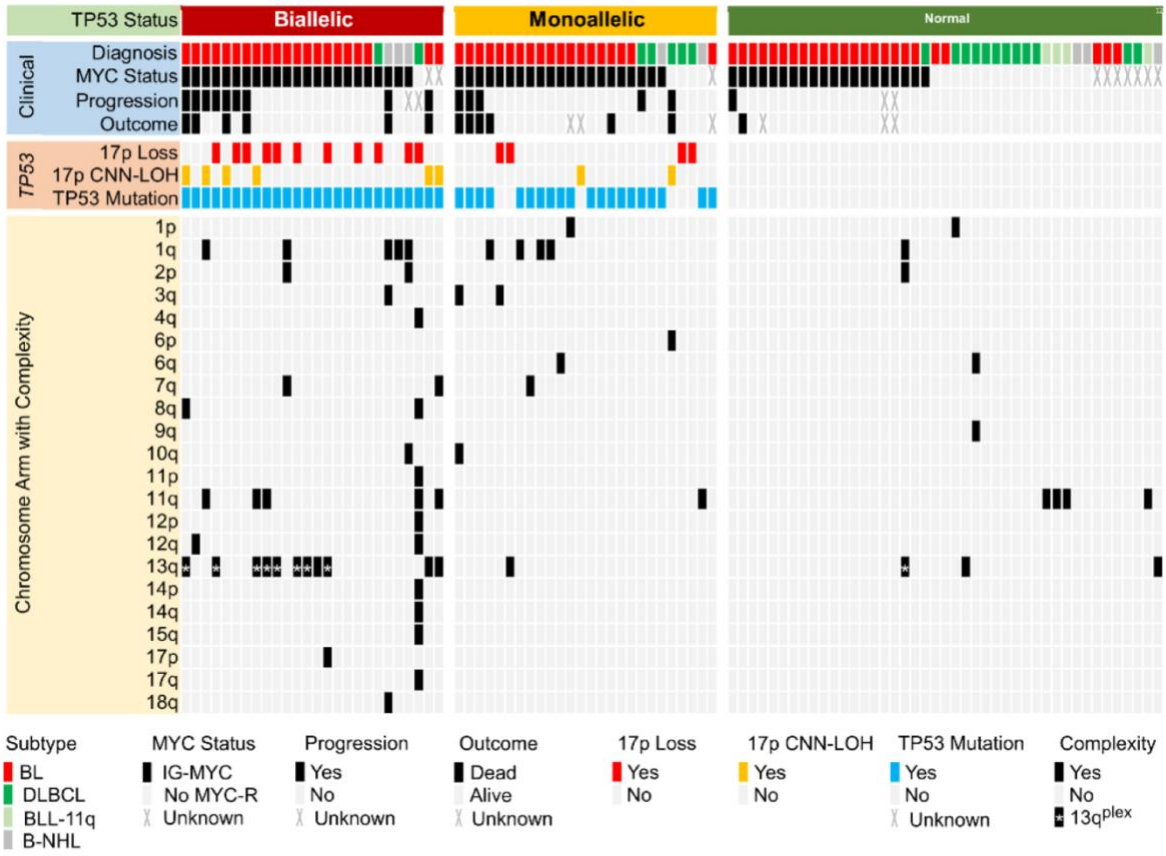
Out of 64 BL cases in the cohort, 40 patients (62.5%) had *TP53* abnormalities (21 biallelic, 19 monoallelic) of which 22 (55.0%) had complex CNA. In contrast, only 1/24 (4.2%) BL cases without a *TP53* abnormality demonstrated complex chromosomal aberrations. Conversely, there was no association between *TP53* abnormalities and complex abnormalities in DLBCL. Five DLBCL cases harboured complex chromosomal aberrations, which only included one monoallelic and one biallelic case. The single biallelic case had Li-Fraumeni syndrome and demonstrated the highest number of complex CNA within the whole cohort with 11 chromosome arms affected.

5.3.10 Complex Chromosomal Aberrations

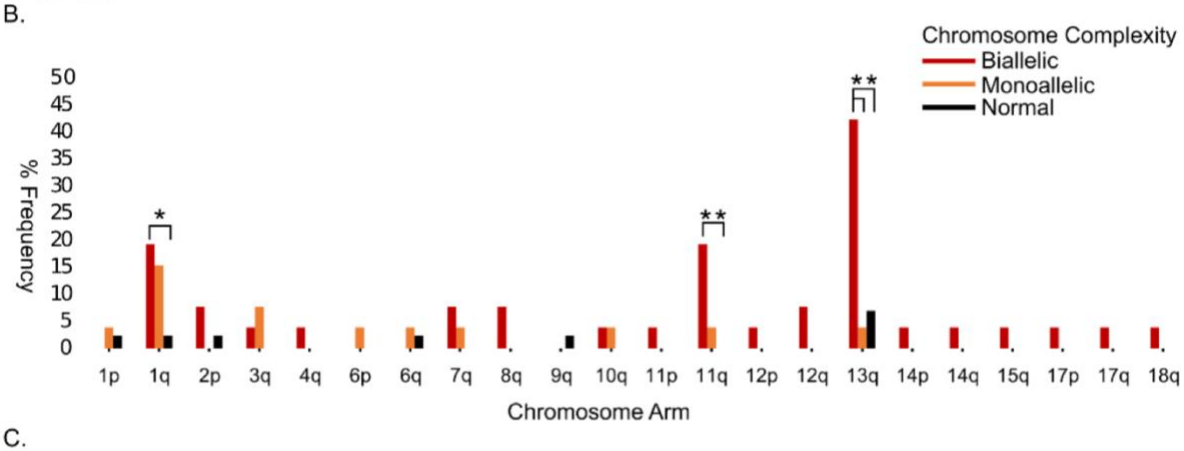
More detailed analysis showed a specific and significant association between *TP53* alterations and complexity of chromosomes 1q, 11q and 13q (Figure 5.8A-C). 66 complex chromosomal abnormalities were identified on 22 different chromosome arms (Figure 5.8B). The most frequently altered arm was 13q, with complex aberrations observed in 15 cases. 1q and 11q abnormalities were both seen in 10 cases. The majority of complex chromosomal abnormalities were found in *MYC*-rearranged BL cases with *TP53* abnormalities. Of 35 cases with chromosomal complexity, 24 (68.6%) were *MYC*-rearranged BL with *TP53* abnormality, compared

to 16/56 (28.6%) *MYC*-rearranged BL cases with *TP53* abnormality and no chromosomal complexity (p<0.001, Fisher exact test).

A.



B.



C.

TP53 Status	1q	11q	13q
Normal frequency	1/43 (2.33%)	0/43 (0%)	3/43 (6.98%)
Monoallelic frequency	4/26 (15.38%)	1/26 (3.85%)	1/26 (3.85%)
Monoallelic vs Normal, *p* value	0.063	0.377	1
Biallelic frequency	5/26 (19.23%)	5/26 (19.23%)	11/26 (42.31%)
Biallelic vs Normal, *p* value	0.026	0.006	0.001

Figure 5.8 – A pattern of associations between complex chromosomal aberrations and TP53 and MYC statuses. (A) An oncoplot showing the association between specific complex chromosomal abnormalities and TP53 abnormalities. (B) Histogram displaying the frequency of complex abnormalities of each chromosome arm in biallelic, monoallelic and TP53 normal groups. As a diagnosis of BLL-11q is determined by the presence of a complex 11q rearrangement, these 4 cases were excluded from this analysis. (C) Summary of the incidence of complex chromosomal

*abnormalities in TP53 biallelic or monoallelic cases compared to TP53 normal cases (Fisher's Exact test *= p<0.05, **=p<0.01).*

5.3.11 Chromosome Arm Abnormalities

13q abnormalities were primarily observed in BL with the exception of one DLBCL case and one B-NHL, NOS case.13/64 (20.3%) BL cases exhibited complex chromosomal aberrations on 13q compared to 1/19 (5.3%) DLBCL cases. The pattern of centromeric gain followed by telomeric loss on 13q resembles the 11q aberration seen in some BL cases in Chapter 4 and also *MYC*-negative Burkitt-like Lymphomas (Salaverria *et al.*, 2014) that were recognised as a provisional subtype of B-NHL in the 2016 WHO Classification of Haematological Malignancies (Swerdlow *et al.*, 2016). Ten cases with this pattern of aberration were found to have complex patterns of gain in addition to the gain-loss pattern, which were termed 13q^{plex} in this thesis, distinguishing them from complex chromosomal abnormalities that did not have the characteristic pattern of gain followed by telomeric deletion on 13q (Figure 5.9A).

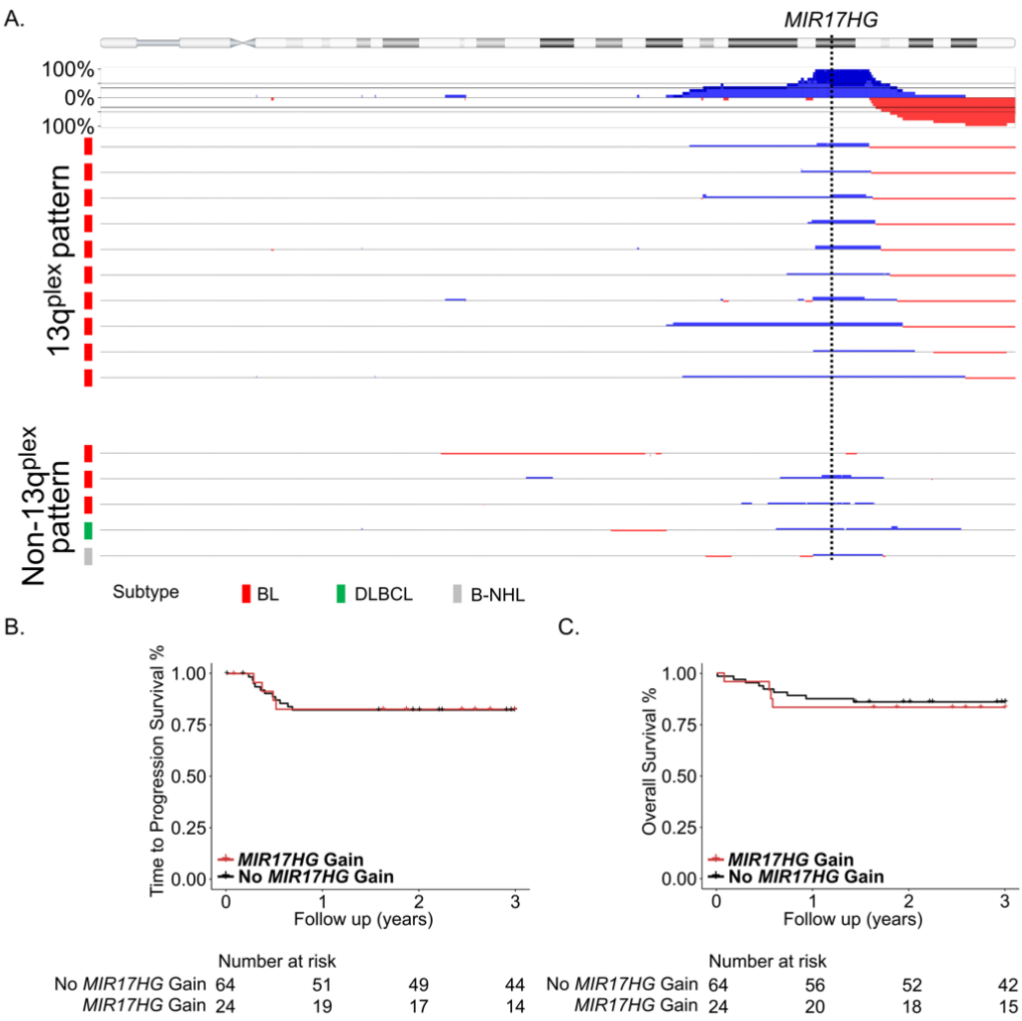


Figure 5.9 - A summary of complex chromosomal aberrations on chromosome arm 13q. (A) Segmented copy number data depicting the size and breakpoint of 13q complexity in ten cases with 13qplex pattern aberrations and five with non-13qplex aberrations. Kaplan Meier plots for MIR17HG gain for (B) time to progression and (C) overall survival

13q^{plex} cases harboured between 1 and 6 distinct segments of copy number gain (median 4). All ten 13q^{plex} cases included amplification of the *MIR17HG* locus, which mapped within the highest peak of copy number gain in 9/10 cases (Figure 5.9A). Telomeric breakpoints across the 13q^{plex} cases, or the transition point between gain and loss, mapped within a 12.1 Mb region, with recurrent breakpoints involving the *HS6ST3* and *MIR4501* loci at 13q32.1. Many of the cases had stepwise copy number gains that did not always correlate with an absolute copy number, suggesting that some of these aberrations are subclonal. This pattern of gain was distinct from

chromothripsis, which is the result of chromosome shattering and manifests as a uniform switching between two copy number states along a chromosome arm. 13q^{plex} aberrations more closely resembled the iAMP21 abnormality seen in paediatric B cell precursor acute lymphoblastic leukaemia (BCP-ALL) (Harewood *et al.*, 2003), which arises due to a breakage-fusion-bridge cycle and is regularly followed by chromothripsis.

The minimally gained region in 13q^{plex} abnormalities harboured the *MIR17HG* microRNA cluster at 13q31.3. Previous published studies have reported an association between gain of *MIR17HG* and relapse (Schiffman *et al.*, 2011) and the cluster has been implicated in the biology of B-NHL in the past (Dal Bo *et al.*, 2015). Amplification of *MIR17HG* was observed more frequently in BL (19/64, 30.0%) than DLBCL (3/19, 15.8%, p=0.375, Fisher exact test) cases, although this difference was not statistically significant. Amplification of 13q31.3 has previously been shown to be associated with paediatric disease by a previous copy number array investigation (Havelange *et al.*, 2016b).

Strikingly, the presence of complexity on 13q had the strongest association with biallelic *TP53* abnormalities (p=0.001, Fisher exact test) (Figure 5.8C), but despite this observation, *MIR17HG* amplification was not associated with disease progression or overall survival in the cohort (Figure 5.9B-C). Whilst the biological relevance of the 13q^{plex} abnormality identified here remains unclear, the association with *TP53* abnormalities may offer an explanation for the postulated prognostic importance of *MIR17HG* amplification in previous smaller cohorts of relapsing BL (Schiffman *et al.*, 2011). Further investigation of 13q abnormalities is required to increase our understanding of how these events occur and whether they have a role in driving disease progression or whether they are simply passenger events; helpful markers of increased genetic instability. Previous studies in other cancers have implicated *TP53* aberration in genomic complexity, and specifically chromothripsis. Sonic-hedgehog medulloblastoma cases with *TP53* aberrations had a high propensity to undergo catastrophic DNA rearrangements such as chromothripsis (Rausch *et al.*, 2012). *TP53* aberrations are also associated with complex karyotypes and specific copy number abnormalities in acute myeloid leukaemia (AML), namely loss of chromosomes 5, 7, 16, or 18 and additional copies of chromosome 1 and 11 (Rucker *et al.*, 2012).

The 11q abnormalities identified in this cohort were mostly typical of those seen in BLL-11q cases (Salaverria *et al.*, 2014) and described in Chapter 4. Two cases had the hallmark gain followed by telomeric loss but had more complex breakpoints between gain and loss with additional aberrations (Figure 5.10).

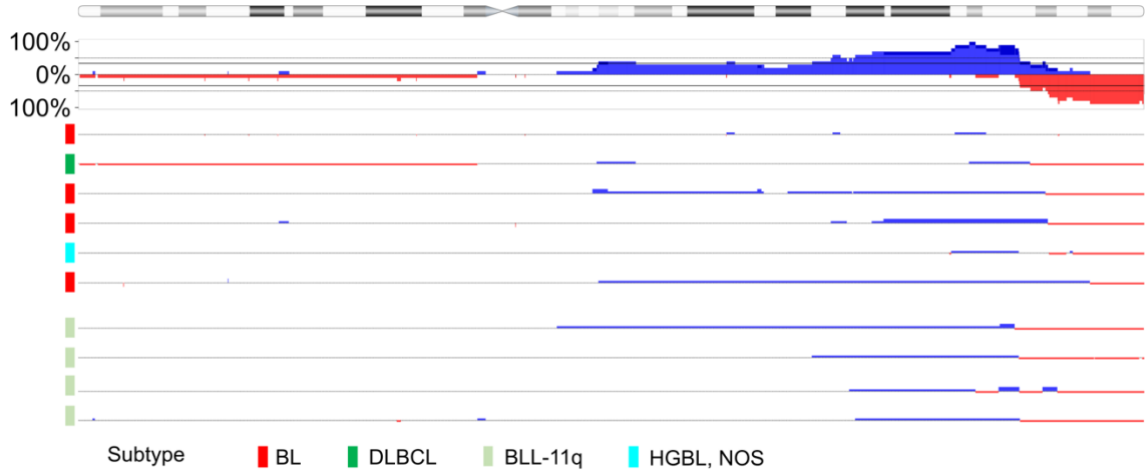


Figure 5.10 - Segmented copy number data showing breakpoints and size of 11q complexity in paediatric B-NHL.

Additionally, one case had a complex pattern of copy number gain but did not harbour the telomeric loss seen in the majority of cases (Figure 5.10). Remarkably, while 11q abnormalities were associated with a biallelic *TP53* status in B-NHL, the four BLL-11q cases had no *TP53* abnormalities. There was a clear difference between breakpoints in BLL-11q cases when compared with BL cases with 11q pattern abnormalities. In BLL-11q cases the breakpoints between gain and loss spanned a 0.6Mb region, while the region in other B-NHL cases was considerably larger (5.2Mb). Further characterisation of 11q abnormalities is presented in Chapter 4.

Chromosome 1q is recurrently gained in aggressive B-NHL cases, including 49/95 cases analysed in this project. Of those, 10 cases had complex copy number gains of 1q (Figure 5.11). Unlike 13q and 11q complexity, deletions on 1q were not as frequent. Instead complex patterns of copy number gain were seen in most cases.

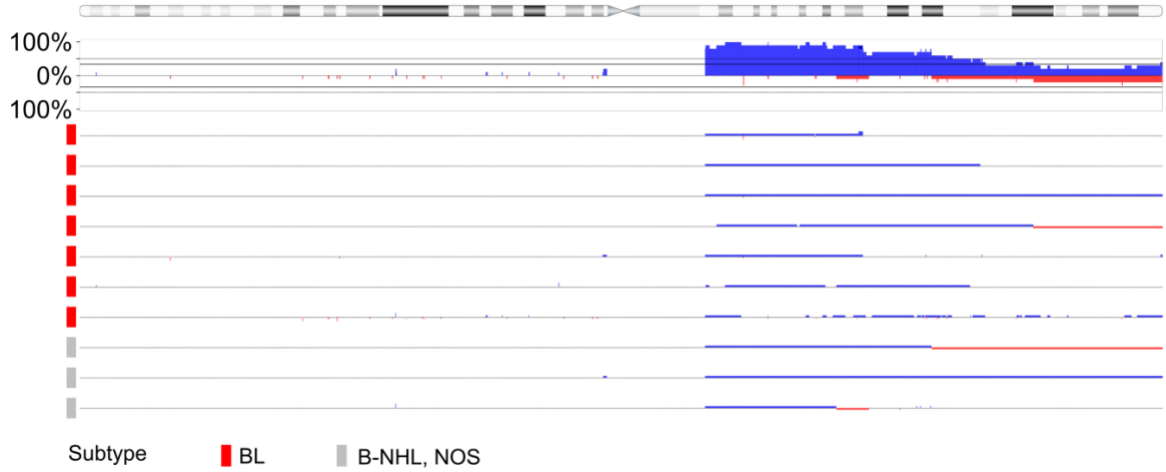


Figure 5.11 - A copy number plot depicting segmented copy number data for cases with 1q complexity.

Complex gain of 1q primarily involved the centromeric region of the arm and typically manifested as fluctuations between copy number states above of three or more. Two cases had telomeric deletions similar to those seen in 13q^{plex} abnormalities in addition to centromeric gain.

One case in the cohort was diagnosed as DLBCL with underlying Li-Fraumeni Syndrome. The extensive complexity observed in the Li-Fraumeni case is further evidence for *TP53* aberration driving genomic instability and complexity in B-NHL (Figure 5.12), as the entire clonal population of the tumour harboured at least one *TP53* abnormality from the initiating clone onwards. Deletion of 17p in this case contributed to the biallelic *TP53* status of the tumour, but it was not known whether this was a somatic event or not.

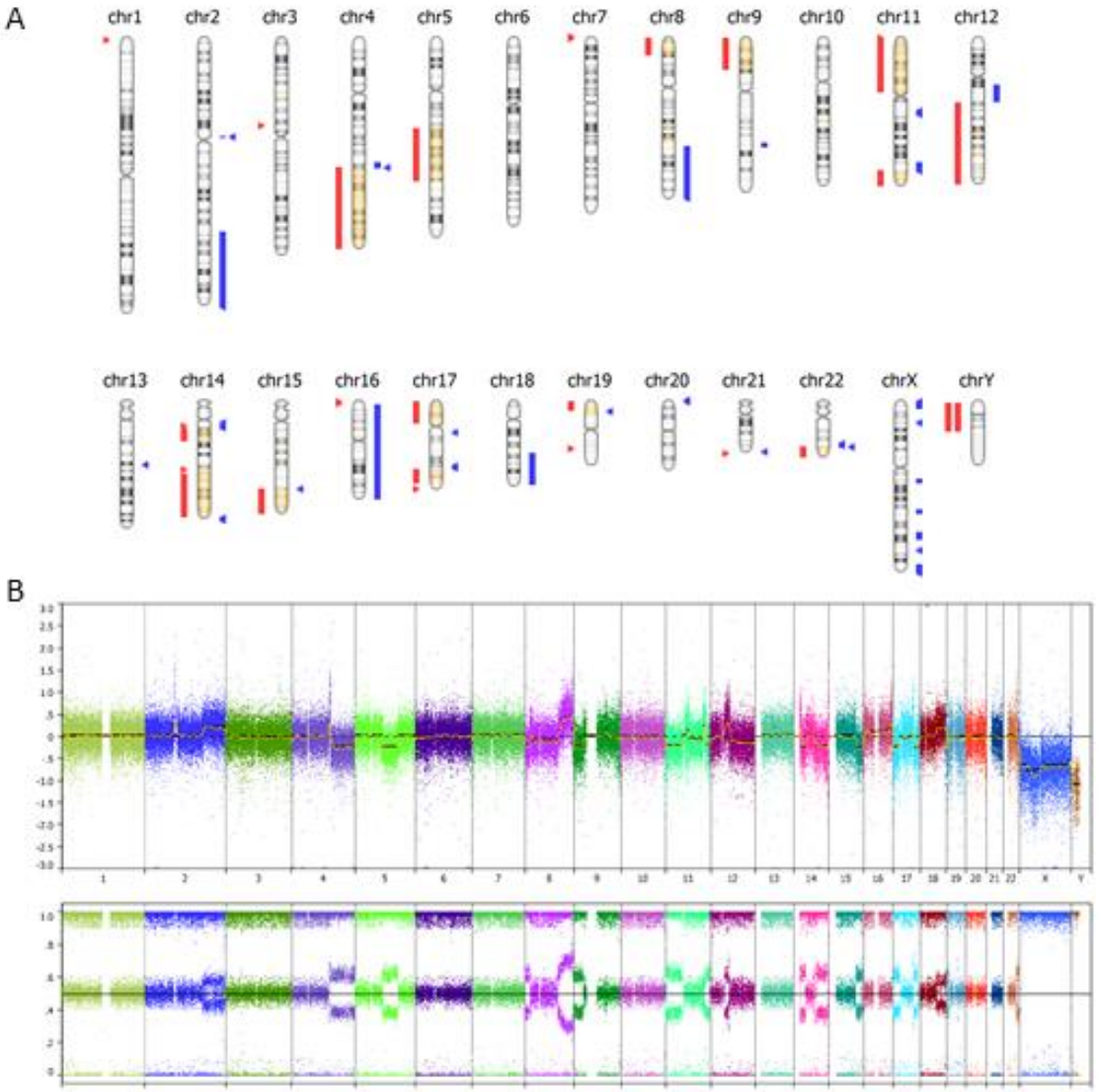


Figure 5.12 - Complex copy number changes observed in a DLBCL tumour from a patient with Li-Fraumeni Syndrome. (A) Ideogram depicting segmented copy number data from the tumour sample. (B) Raw log ratio and B-allele frequency data depicting complex aberrations identified in the tumour.

5.4 Discussion

Genomic analyses of BL and DLBCL have reported that *TP53* mutations are among the most recurrent mutations in B-NHL (Giulino-Roth *et al.*, 2012; Love *et al.*, 2012) and that 17p copy number abnormalities are recurrent (Poirel *et al.*, 2009; Scholtysik *et al.*, 2010; Schiffman *et al.*, 2011; Scholtysik *et al.*, 2015) but the clinical relevance and impact on patient outcome has not previously been elucidated.

This study has identified, for the first time, an association between *TP53* abnormalities and both disease progression and poor overall survival for UK children diagnosed with mature B cell non-Hodgkin lymphoma. We showed that over half of paediatric B-NHL patients harboured at least one abnormality affecting the *TP53*, and that those harbouring aberrations on both alleles were at the highest risk of disease progression and death. The exciting and novel findings of this investigation have the potential to improve outcome for B-NHL patients as we present a prognostic genomic marker that is strongly associated with disease progression that can be easily screened for in the clinic.

Equally remarkable is the finding that patients with normal *TP53* status at diagnosis did exceedingly well on standard frontline therapy. Only one patient out of 40 without *TP53* aberrations at diagnosis went on to develop progressive disease, and this case went on to acquire clonal *TP53* abnormalities by the point of disease progression. Using *TP53* status to stratify patients highlights a subset of patients who could benefit from treatment de-escalation. There is precedent in the literature for integrating *TP53* mutation and copy number status with treatment strategy in the setting of in CLL in a practical way (Campo *et al.*, 2018). The foundations for comprehensive investigation of *TP53* status already exist in haematological laboratories across Europe as a result of the *TP53* Network implemented by the European Research Initiative on CLL (ERIC) (Malcikova *et al.*, 2018).

It has been hypothesised that inhibition of p53 protein is functionally required in all BL cases to bypass pro-apoptotic signals from deregulated MYC expression, which is a common mechanism in other MYC+ tumours (Eischen *et al.*, 1999). High expression of MDM4 has previously been demonstrated in *TP53*wt BL suggesting that

abrogation of the p53 pathway is an important requirement for BL development (Leventaki *et al.*, 2012).

Recent studies and trials have assessed and suggested the therapeutic potential of MDM2/MDMX inhibitors in paediatric cancer and haematological malignancies, including B-NHL (Tisato *et al.*, 2017). The work detailed in this study casts doubt on the relevance of these inhibitors in the treatment of relapsed/refractory paediatric B-NHL, as the cases with the worst outcome on current clinical protocols are almost all *TP53* abnormal. MDM2/MDMX inhibition is still a promising candidate for the treatment of B-NHL in the *TP53* normal group to reduce the intensity of chemotherapy in low-risk patients. This is further supportive evidence for ascertaining the *TP53* status of paediatric B-NHL at the point of diagnosis.

Genome complexity associated with *TP53* abnormality in paediatric B-NHL has not been described previously. This association was strongest in *MYC*-rearranged BL patients with *TP53* aberration. Despite the association between genomic complexity and *TP53* abnormalities, complexity was not in itself associated with poor outcome. Chromosome arms 13q, 11q and 1q were most frequently affected and are known to be the sites of recurrent non-complex copy number alterations in paediatric B-NHL (Scholtysik *et al.*, 2010; Schiffman *et al.*, 2011; Salaverria *et al.*, 2014; Scholtysik *et al.*, 2015; Havelange *et al.*, 2016b). It is noteworthy that non-complex abnormalities of 13q and 1q were more frequent in the cohort but were not associated with *TP53* status. *MIR17HG* amplification was a common factor of all but one complex 13q aberration, and all $13q^{plex}$ pattern cases. While it was determined that *MIR17HG* amplification was not a marker of poor outcome in this cohort, the strong association between $13q^{plex}$ pattern aberrations and *TP53* abnormality, as well as its established biological role in B-NHL, may explain why previous studies have suggested *MIR17HG* as a prognostic marker. Both 1q and 11q harbour p53 pathway-associated genes which may synergise with *TP53* aberrations to aid lymphoma development and progression. However it is very likely that a large proportion of complex chromosomal aberrations identified in this cohort were passenger abnormalities and did not contribute to the pathogenesis of the malignancy.

TP53 status is the first clinically actionable genomic prognostic marker for paediatric B-NHL patients. Patients could be routinely screened for aberration using an existing

infrastructure within UK hospitals. The next step is to validate the findings of this study in a large, clinically annotated cohort from a prospective trial to facilitate translating these findings to the clinic. The results of this study have the potential to advance the development of more efficacious and less toxic treatment options for patients. Diagnostic *TP53* status testing will improve our ability to stratify patients in a clinically meaningful way and improve their outcome during treatment and beyond.

The significance of *TP53* aberrations is still not understood in other BL subtypes, particularly endemic BL. A recent study by Grande *et al.* (2019) identified a significant difference between subgroups in the frequency of *TP53* mutations based on EBV status, with EBV negative BL having twice the rate of *TP53* mutation compared to EBV positive cases. Despite recent improvements in survival rates of paediatric endemic BL at the Queen Elizabeth Central Hospital in Blantyre, Malawi, cure rates still lag behind those seen in the UK. Effective prognostic markers of disease progression are urgently needed to improve outcomes for patients without increasing treatment toxicity. Chapter 6 of this thesis details copy number analysis of endemic and sporadic BL patients with a view to identify prognostic markers.

Chapter 6: Comparison of Genomic Landscapes in Endemic BL and Sporadic BL

6.1 Introduction

Endemic Burkitt lymphoma represents the first subtype of BL that was characterised. Initially described in 1958 by Denis Burkitt in Africa, endemic BL frequently presents as swelling of the jaw or eye socket (Burkitt, 1958). The disease is the most common childhood cancer in sub-Saharan Africa, found in the warmer, wetter regions of the continent. The incidence of endemic BL was found to be consistent with the occurrence of the malarial parasite *Plasmodium falciparum*, leading to a longstanding belief that the parasite was instrumental in the pathogenesis of BL. Investigations into BL shortly later revealed the presence of a novel oncogenic virus, the Epstein-Barr virus (Epstein *et al.*, 1964). It is believed that both play a role in disease through inducing AID (activation induced cytidine deaminase) (Figure 6.1).

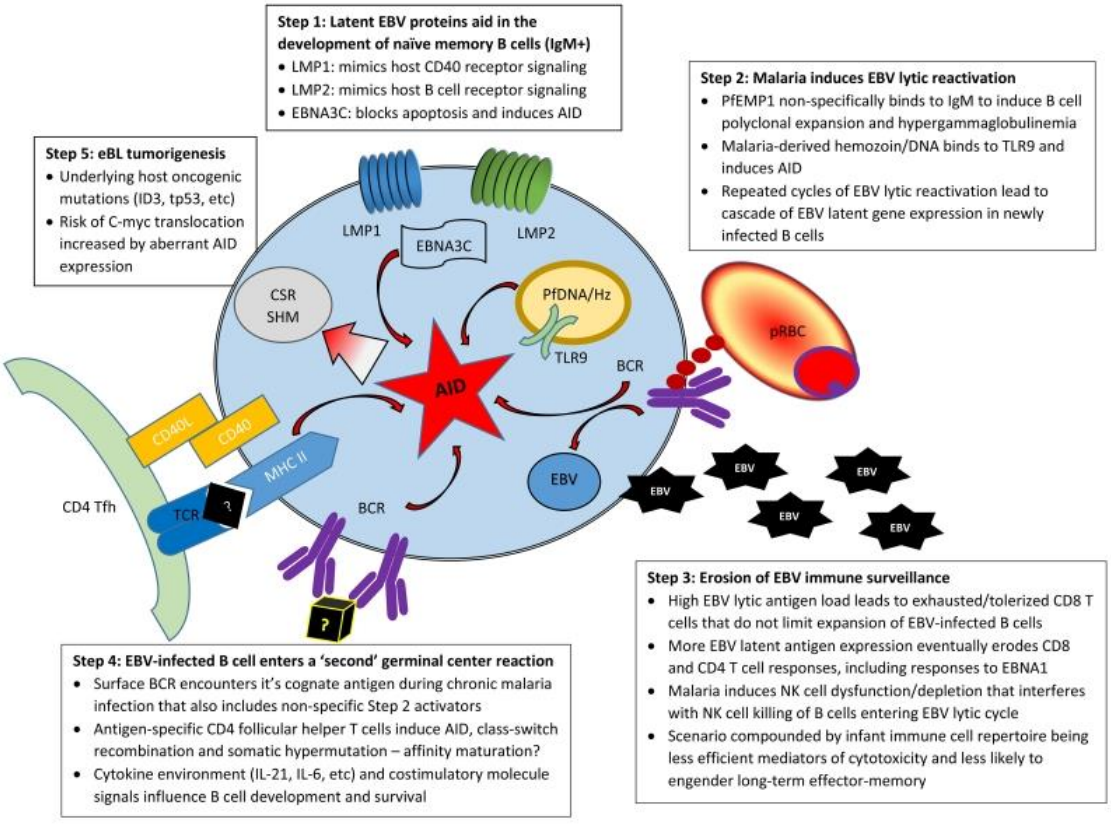


Figure 6.1 – Proposed synergistic mechanisms by which Malarial infection may contribute to increased AID expression in an EBV infected B cell in Burkitt Lymphoma development. Adapted from (Moormann and Bailey, 2016)

Patients diagnosed with endemic Burkitt Lymphoma generally have a poorer prognosis than those diagnosed with the sporadic subtype in the UK and there are a

number of factors contributing to this inferior outcome. Patients treated at the Queen Elizabeth Central Hospital, Blantyre, Malawi are given a significantly reduced treatment protocol compared to patients in the UK. Several protocols exist in different clinical and geographical settings and there is little consistency in these protocols, resulting in event-free survival rates ranging from 50-80%. FAB/LMB96-based chemotherapy administered to patients in the UK comes with severe toxicity and requires a high degree of supportive care for months. Reduced-intensity therapy in lower income countries lowers supportive care needs for the patient, reducing the likelihood of severe adverse events, as well as the cost of treatment per patient (Hesseling *et al.*, 2003).

We hypothesised that while the highly chemotherapy-resistant BL cases have poor outcomes in a sporadic setting still occur in the endemic subtype, they are currently indistinguishable from a larger population of patients who are not sufficiently treated by endemic protocols. Intensifying therapy for all patients is not feasible due to the cost of chemotherapeutic agents and supportive care (Hesseling *et al.*, 2008), however if patients with poorer risk could be identified at diagnosis then an intensified regimen could be administered to patients who need it. Previous studies (Poirel *et al.*, 2009; Schiffman *et al.*, 2011) and work in chapters 4 and 5 have identified prognostic genomic markers that confer a poorer outcome in sporadic BL and B-NHL. If these or other genomic markers can be used in the endemic setting then the outcome of patients with endemic BL could be significantly improved by more targeted treatment.

Independently of the treatment prescribed to the different subtypes, it is well understood that sporadic and endemic are biologically distinct (Ferry, 2006). The genetic profiles of both subtypes share many features, such as the hallmark IG/*MYC* translocation and mutations of the ID3, TCF3 and CCDN3 axis, but distinct differences have been described in the incidence of such mutations (Abate *et al.*, 2015). *IGH-MYC* translocation breakpoints are different between subtypes, with the majority of rearrangements in sporadic BL truncating the first exon while endemic cases harbour rearrangements adjacent to the *MYC* gene (Pelicci *et al.*, 1986). Abate *et al.* (2015) performed variant calling on RNA-Seq data from 20 endemic BL cases and identified *TP53* mutations at a lower incidence in endemic BL (3/20, 15.0%) than in the sporadic form (14/41, 34.1%). This finding is significant given the finding reported in Chapter 5 that *TP53* aberrations are associated with disease progression

and poor outcome in paediatric B-NHL treated in a UK setting. Conversely, *ARID1A* mutations were shown to be more common in endemic BL (Abate *et al.*, 2015) leading to a loss of expression, a finding which has also been reported in other EBV positive cancers (Abe *et al.*, 2012).

It has been repeatedly reported that differences in genomic profiles between subtypes is thought to driven by EBV status, rather than geographical factors (Abate *et al.*, 2015; Grande *et al.*, 2019). Over 90% of endemic BL cases are EBV+ while the number is under 30% in the sporadic subtype (Ferry, 2006) and even as low as 5% in some studies (Zhou *et al.*, 2019). A recent study found that there were a significantly higher number of mutations in the EBV+ cases, primarily endemic BL, but that they featured fewer driver mutations (Grande *et al.*, 2019). This was attributed to a lower incidence of *SMARCA4* and *CCND3* mutations in EBV+ cases as well as fewer apoptosis-associated gene mutations, including *TP53* and *USP7*. The authors of the study postulated that *EBNA1* expression, driven by EBV+ infection, compensates for the anti-apoptotic signal that these mutations would provide. Non-coding mutations were identified at a significantly higher rate in EBV positive BL cases and this was shown to be associated with higher AID activity and aberrant somatic hypermutation (Grande *et al.*, 2019). A more recent whole genome sequencing study of BL subtypes also observed a higher mutational load in EBV positive BL that was attributed to AID-associated mutations (Panea *et al.*, 2019).

FOXO1 is a commonly mutated gene in DLBCL (Trinh *et al.*, 2013) and BL (Kabrani *et al.*, 2018) but recent work in our lab has shown that the rate of mutation between endemic and sporadic subtypes are significantly different (Zhou *et al.*, 2019). Mutations in *FOXO1* are commonly found within a hotspot in the AKT recognition motif. However, mutations were found at different residues within the hotspot depending on the subtype.

6.2 Chapter Aims

The specific aims of the work detailed in this chapter were as follows:

- Analyse copy number data from eBL and sBL cohorts to identify recurrent copy number analysis using integrative bioinformatics.
- Evaluate the prognostic significance of recurrent abnormalities in the endemic cohort.

- Investigate *TP53* abnormalities in eBL and compare findings with those in Chapter 5, determining whether *TP53* status is clinically important in both BL subtypes.

6.2 Results

6.3.1 Patient cohorts and demographics

Fine needle aspirates (FNA) from paediatric patients diagnosed at the Queen Elizabeth Central Hospital (QECH), Blantyre, Malawi were obtained and reviewed by pathologists at the Royal Victoria Infirmiary in Newcastle, UK. Of these, 245 were determined to be BL after diagnostic pathology was performed on cases with sufficient material. Affymetrix Oncoscan v2 MIP arrays were initially generated for a core set of 60 cases (Batch 1), with a second batch of 45 arrays generated to give a total of 105 cases with copy number data. Endemic BL patients were treated with one of four protocols described in Figure 6.2.

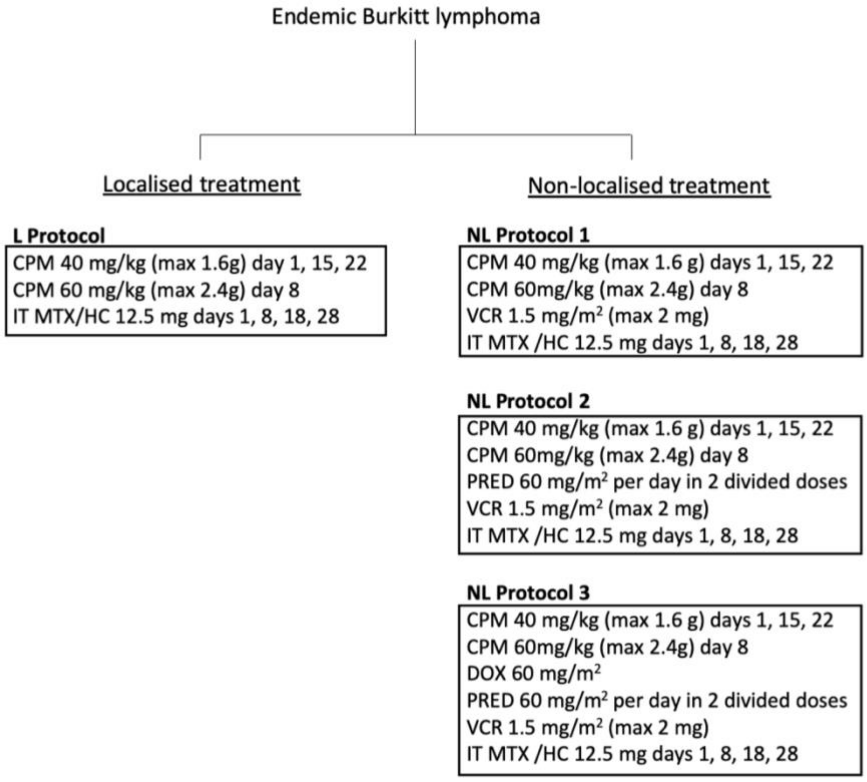


Figure 6.2 - Treatment protocols used at frontline therapy for endemic BL patients treated at the Queen Elizabeth Central Hospital, Blantyre, Malawi.

The addition of doxorubicin brought the existing endemic BL protocol closer to those used in the UK with superior overall survival and lower risk of relapse (Miles *et al.*, 2012). EBER status was known for 51/105 (48.6%) patients and all patients screened were EBV positive. EBER status was comprehensive for batch 1, with 51/60 (85.0%) patients screened. Data for Batch 2 was mostly unavailable; one case was screened and was found to be EBV positive.

The sporadic BL cohort analysed in this study involved all 75 BL cases acquired from the CCLG biobank that were analysed in chapters 4 and 5. An additional four samples were added from patients treated at the Great North Children's Hospital. DNA was extracted from viable cells. Three additional BL cases were identified from the CCLG cohort that did not have material available when the original copy number study was performed.

6.3.2 Segmentation and Quality Control of Copy Number Data from SNP Arrays

Copy number data from Oncoscan arrays for 105 eBL cases was analysed using Nexus Copy Number 10.0. Compared to Oncoscan array data generated from the sporadic CCLG cohort the data generated from Malawi cases had more intra-assay variability in log ratio values. There were 170 frequently seen copy number gains called by the segmentation algorithm that were recurrent in the cohort regardless of when the array was hybridised. The false positive segments of copy number gain were characterised by non-distinct breakpoints and no discernible change in the BAF track. The regions were identified manually, and a custom BED file was used to filter the region out of analysis within Nexus Copy Number 10.0 and GISTIC2.0.

"Waviness" in aCGH and copy number/SNP array data is a well-known occurrence and source of noise (Lepretre *et al.*, 2010). Systematic correction performed using Nexus Copy Number 10.0 corrects for this array-specific pattern of noise. Despite this waviness was still observed in some regions for the majority of samples. This coincided with the majority of the additional regions of gain that were filtered out. Sporadic BL cases did not exhibit waviness or recurrent gains seen in the eBL cohort. Noise identified in sBL cases was characterised previously in chapter four, and the CNV filter file used in GISTIC2.0 analysis was re-used in this study.

In endemic cases the mean number of copy number events and CNN-LOH events was 58.4 (range 0-331) and 0.8 (range 0-4) respectively. The mean number of copy number aberrations and CNN-LOH events in sporadic BL was 71.2 (range 0-761) and 1.6 (range 0-20) respectively. Segmented copy number data for both subtypes is displayed in Figure 6.3.

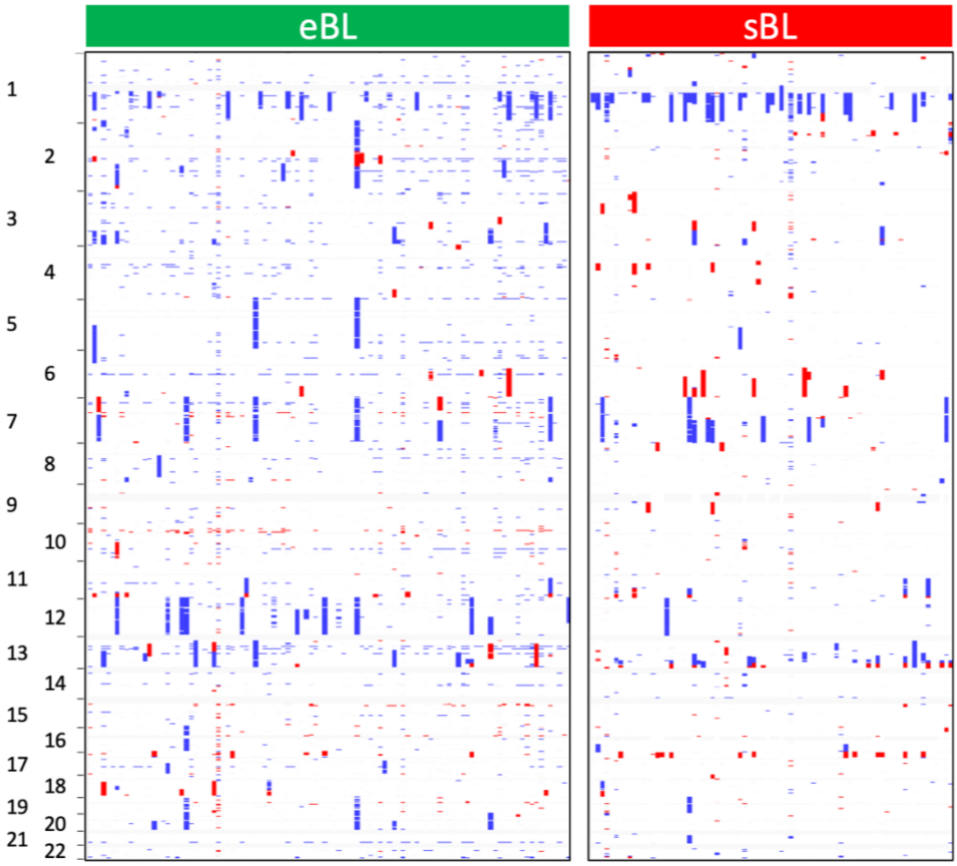


Figure 6.3 - A heatmap showing segmented copy number data for 105 endemic and 79 sporadic Burkitt Lymphoma. Blue denotes gain and red denotes loss.

6.3.3 GISTIC2.0 Analysis of Significant Targets in Diagnostic eBL and sBL Samples

GISTIC2.0 analysis was performed separately on 105 eBL and 79 sporadic BL patients to identify the most frequent copy number aberrations in both subtypes. CNN-LOH was also run through GISTIC2.0 to determine frequently affected chromosome arms which may be sites of recurrent driver mutations. Results for both cohorts are summarised in Table 6.1 and Table 6.2.

cytoband	Event	q value	residual q value	wide peak boundaries
3q29	Gain	<0.0001	<0.0001	chr3:195331537-198022430
1q21.1-1q21.3	Gain	0.035982	0.035982	chr1:145734732-150974161
1q31.2-1q31.3	Gain	0.035982	0.035982	chr1:191709542-193922252
2q34	Gain	0.035982	0.035982	chr2:212206138-212571693
22q13.2-22q13.31	Gain	0.035982	0.035982	chr22:41375250-44298893
1p34.3	Gain	0.037794	0.037794	chr1:35085667-36767600
5q12.3	Gain	0.037794	0.037794	chr5:64670271-66311589
2q24.1-2q33.2	Gain	0.035982	0.039698	chr2:155305838-204450312
9q34.3	Gain	0.050387	0.050387	chr9:138189380-139554402
17q11.2	Gain	0.066473	0.066473	chr17:26880986-28944928
13q31.1-13q32.2	Gain	0.07831	0.07831	chr13:86730661-98383041
2p21	Gain	0.090581	0.090581	chr2:42651748-46242712
3p21.2-3p21.1	Gain	0.091593	0.091593	chr3:51370934-52705478
11q23.3-11q25	Loss	<0.0001	<0.0001	chr11:120850978-135006516
2p24.3-2p24.2	Loss	<0.0001	<0.0001	chr2:15695921-16738751
17p12	Loss	<0.0001	<0.0001	chr17:10741335-11886863
2q13-2q24.1	Loss	0.0049636	0.0049636	chr2:114253448-157186561
18q21.31-18q23	Loss	0.0087277	0.0088288	chr18:55462713-78077248
1q43-1q44	Loss	0.058405	0.058405	chr1:241676066-249250621
13q12.3-13q31.1	Loss	0.058405	0.058405	chr13:30521044-86372843
10p11.22-10q23.31	Loss	0.068386	0.068386	chr10:33623137-91458197
1p36.33-1p36.11	CNN-LOH	<0.0001	<0.0001	chr1:1-25226220
17q24.2-17q25.3	CNN-LOH	0.00010378	0.00010377	chr17:65042756-81195210
Chromosome 1	CNN-LOH	<0.0001	1	chr1:1-249250621

Table 6.1 - Focal regions identified by GISTIC2.0 analysis in eBL.

cytoband	Event	q value	residual q value	wide peak boundaries
13q31.3	Gain	<0.0001	<0.0001	chr13:91686078-94356987
1q23.1	Gain	0.0059876	0.0059876	chr1:157410003-157601372
3q29	Gain	0.02211	0.02211	chr3:195539955-198022430
9p21.3	Loss	<0.0001	<0.0001	chr9:21860429-22448736
13q33.3-13q34	Loss	<0.0001	<0.0001	chr13:108512716-115169878
3q13.31-3q13.32	Loss	<0.0001	<0.0001	chr3:116160574-118627418
2p22.2-2p16.3	Loss	0.00038927	0.00042946	chr2:36753571-50591528
4q13.3-4q22.1	Loss	0.0015957	0.0015957	chr4:62351341-89645650
1p36.12-1p36.11	Loss	0.0074315	0.0074164	chr1:23733994-23955103
6q22.33-6q27	Loss	0.038012	0.04168	chr6:129880109-171115067
6q13-6q21	Loss	0.044396	0.047302	chr6:73316478-109641966
3q26.31-3q26.33	Loss	0.084724	0.079414	chr3:175520336-181332410
1p36.33-1p34.3	CNN-LOH	0.0017775	0.0017775	chr1:704575-36603388
Chromosome 17	CNN-LOH	0.0094264	0.0094264	chr17:1-81195210
1q25.2-1q32.1	CNN-LOH	0.075857	0.075617	chr1:179880905-204174327

Table 6.2 - Focal regions identified by GISTIC2.0 analysis in sBL.

In total, 27 regions were identified in the eBL cohort by focal GISTIC2.0 analysis. After manual investigation of the regions, three were excluded for being likely noise/CNV regions: 17q11.2 gain, 8p23.2 gain and 15q11.2 loss (Figure 6.2A-B). A total of 35 regions were highlighted in the sBL cohort, which were reduced to 15 after manual checks (Figure 6.4A-B).

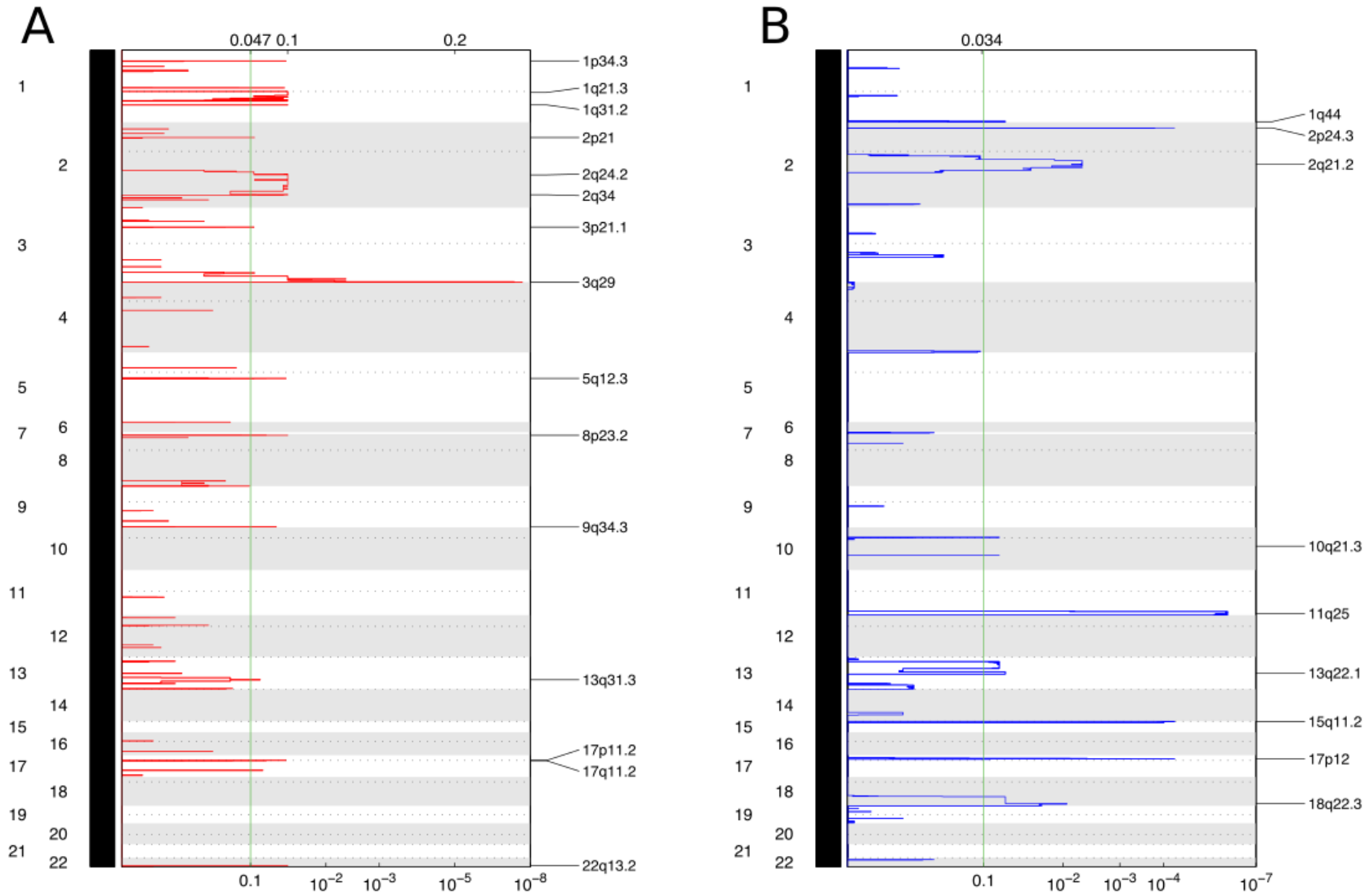


Figure 6.4 - GISTIC2.0 plots showing recurrent regions of (A) copy number gain and (B) deletion in eBL.

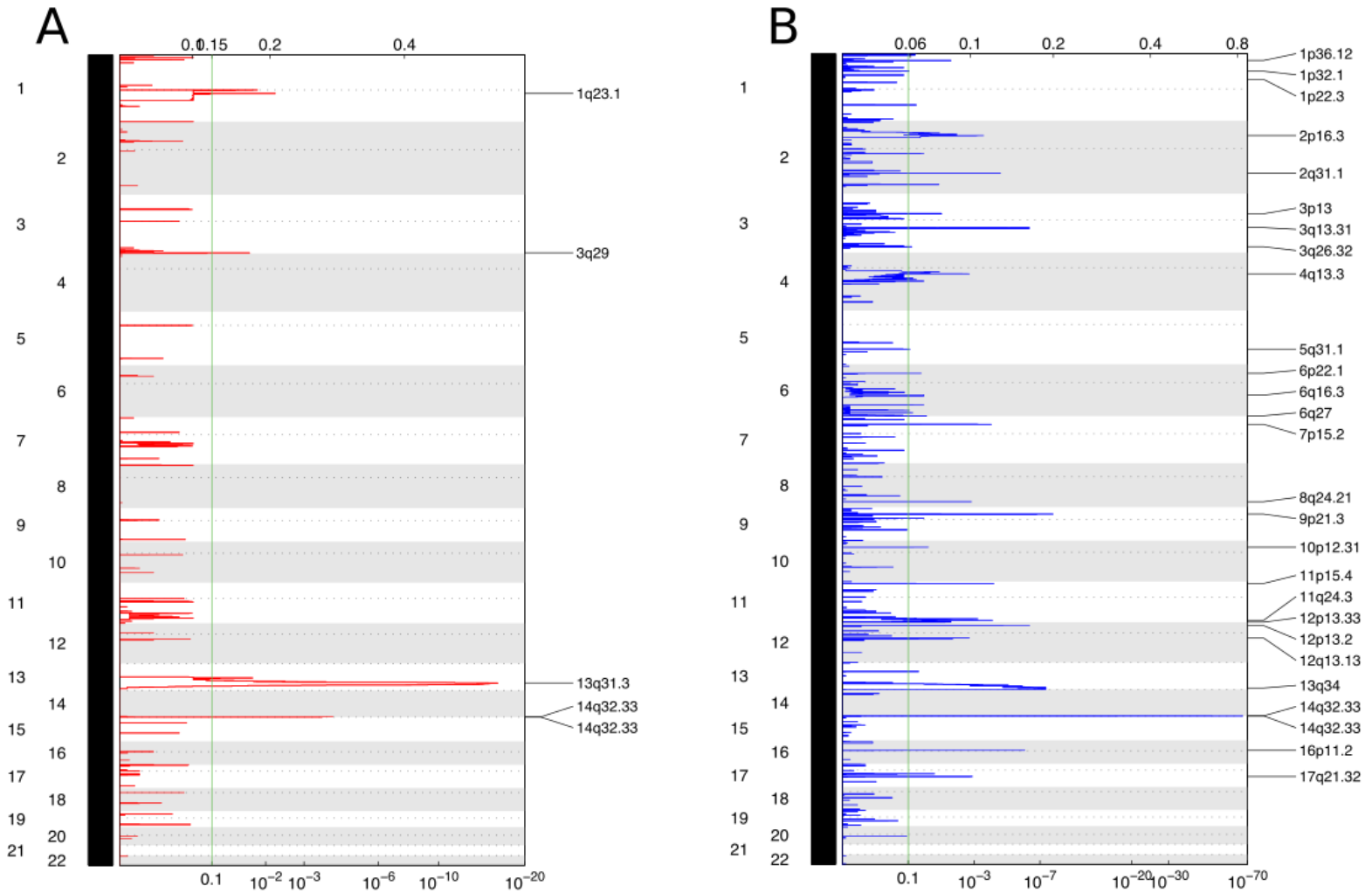


Figure 6.5 - GISTIC2.0 plots showing recurrent regions of (A) copy number gain and (B) deletion in sBL.The following regions were identified by GISTIC2.0 but were excluded after manual analysis: Two separate peaks of copy number gain at 14q32.33 and copy number loss at 1p32.1, 1p22.3, 2q31.1, 3p13, 5q31.1, 6p22.1, 7p15.2, 8q24.21, 10p21.31, 11p15.4, 11q24.3, 12p13.33, 12p13.2, 12q13.13, 14q21.22, 14q32.33, 16p11.2 and 17q21.32.

Regions of copy number gains were more prevalent in the eBL cohort (13/24 regions identified in eBL, 54.2% vs 3/15 regions identified in sBL 20.0%, p=0.049, Fisher exact test), while CNN-LOH was infrequently observed in both cohorts. Only three recurrent regions of focal copy number gain were observed in the sBL cohort; 13q31.3, 1q23.1 and 3q29 (Figure 6.6).

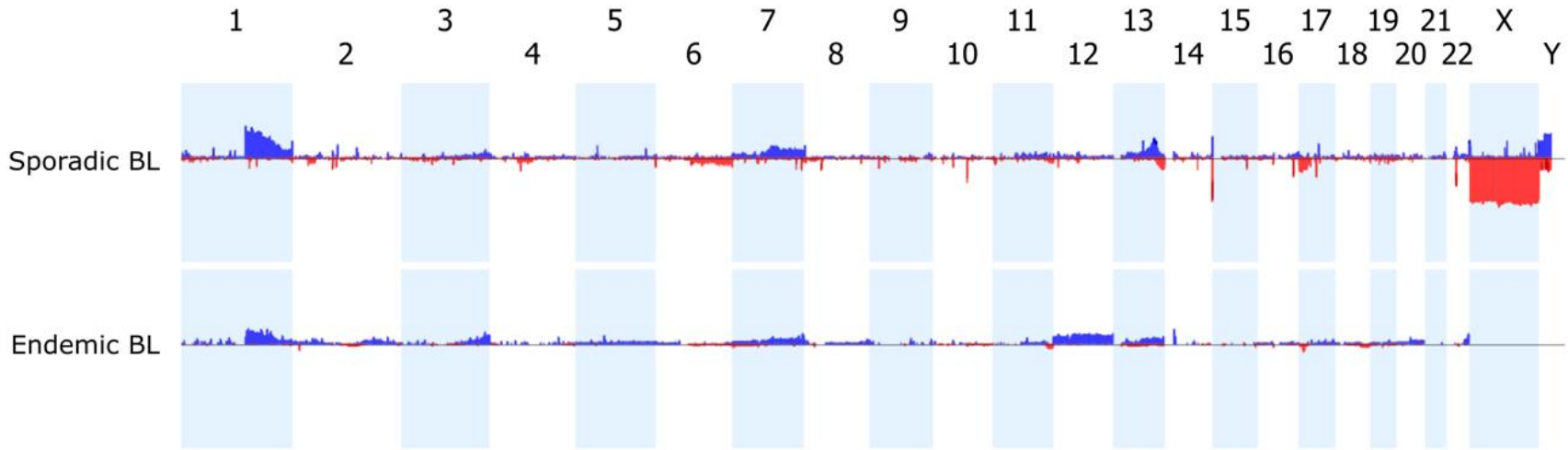


Figure 6.6 - Copy number aggregate plot showing the copy number profiles of sBL and eBL analysed by copy number array. Blue peaks denote copy number gain and red peaks denote copy number loss.

Figure 6.6 shows that the most frequently observed regions of copy number gain were on 1q and 3q29 in both cohorts (Table 6.1, Table 6.2, Figure 6.6). 1q gain is reported as the most frequent genomic aberration in BL after *MYC* rearrangement (Poirel *et al.*, 2009) and the minimal common region was adjacent to the centromere in each subgroup. 24/105 (22.9%) eBL cases had gain of 1q compared to 35/79 (44.3%) sBL cases (p=0.002, Fisher exact test) (Figure 6.6). In sBL the region most frequently gained was 1q23.1 while in eBL 1q21.1-1q21.3 was most prevalent. 3q29 amplification was commonly seen in both cohorts (Figure 6.6) and represents a region of interest in paediatric B-NHL as a prognostic marker identified in Chapter 4. The minimal common region of gain on 3q29 was similar in both cohorts, involving to the telomere in both cases. Chromosome 13 gain involving the *MIR17HG* miRNA cluster at 13q31.3 was seen in both cohorts but was significantly more frequent in sBL (10/105, 9.5% eBL and 21/79, 26.6% sBL, p=0.003, Fisher exact test). The minimal common region of copy number gain at 13q31.3 was larger in the eBL cases at 11.7Mb compared to 2.7Mb in sBL. Loss of 6q frequently observed in sporadic BL but was seen in only one patient with eBL and additional copies of chromosome 12 were almost exclusively seen in eBL. 17p deletion involving *TP53* was present at a similar frequency in eBL and sBL (16/105, 15.2% eBL vs 12/79, 15.2% sBL, p=1, Fisher exact test).

6.3.4 Abnormalities associated with disease progression in eBL

Relapse was more common in endemic BL than sBL. 27 patients experienced relapse disease while 58 were known to not experience any progression. Outcome data was unavailable for 20 eBL cases. Risk of relapse (RR) analysis included only patients who completed front-line therapy and time to relapse was measured from date of diagnosis to date of relapse. Patients that did not have a documented relapse event but went on to die, either from disease or other causes, were excluded.

		sporadic BL	endemic BL
Total Cases		79	105
Median Age at Diagnosis (range), years		8 (0-19)	7 (2-16)
Sex	Male	61 (77.2%)	60 (57.1%)
	Female	13 (16.5%)	43 (41.0%)
	U	5 (6.3%)	2 (1.9%)
Tumour Stage	High (3 or 4)	46 (58.2%)	73 (69.5%)
	Low (1 or 2)	8 (10.1%)	23 (21.9%)
	U	25 (31.6%)	9 (8.6%)
CSF Involvement	Y	2 (2.5%)	9 (8.6%)
	N	50 (63.3%)	70 (66.7%)
	U	27 (34.2%)	26 (24.8%)
BM Involvement	Y	17 (21.5%)	11 (10.5%)
	N	44 (55.7%)	59 (56.2%)
	U	18 (10.1%)	35 (33.3%)
CNS Involvement	Y	5 (6.3%)	ND
	N	53 (67.1%)	-
	U	21 (26.6%)	-
EBV Positive	Y	2 (2.5%)	51 (48.6%)
	N	39 (49.4%)	0 (0.0%)
	U	38 (48.1%)	54 (51.4%)
MYC translocation	Y	70 (88.6%)	105 (100.0%)
	N	5 (6.3%)	0 (0.0%)
	U	4 (5.1%)	0 (0.0%)
Treatment	FAB/LMB 96 Protocol	62 (78.5%)	-
	Group A	2 (2.5%)	-
	Group B	42 (53.2%)	-
	Group C	13 (16.5%)	-
	Group unknown	22 (27.8%)	-
	Malawi 28 day protocol†	-	88 (83.8%)
	with doxorubicin	-	76 (72.4%)
	reduced protocol	-	6 (5.7%)
	No Treatment	-	0 (0.0%)
	U	17 (21.5%)	13 (12.4%)
Follow-up	Number of cases with follow-up data	72 (91.1%)	85 (81.0%)
Disease progression	Relapse/refractory disease	16 (20.3%)	27 (25.7%)
	Median time to relapse (range), months	4.8 (2.8-6.7)	5.3 (1.6-10.9)
Overall survival	Dead	15 (19.0%)	28 (26.7%)
	Median time to death (range), months	5.3 (0.1-11.1)	3.8 (0-14.1)

Table 6.3 - Clinical characteristics of sBL and eBL cohorts. ND= Not done.

A copy number aggregate plot comparing the genomic profiles of cases that progressed and did not progress is shown in Figure 6.7. Abnormalities in regions identified by GISTIC2.0 analysis occurred in both patient subgroups.

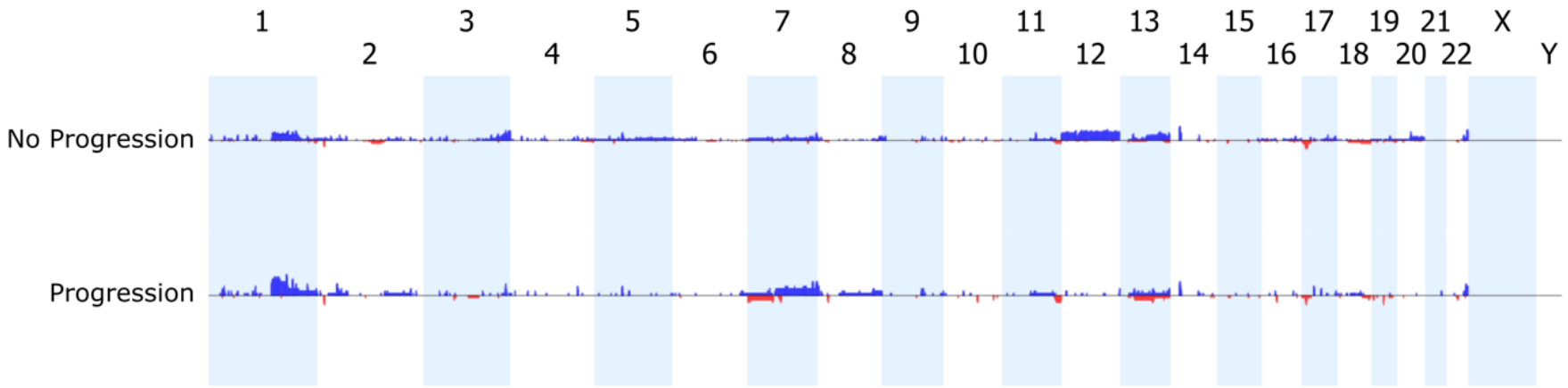


Figure 6.7 - A copy number aggregate plot showing the copy number profile at diagnosis of cases with and without disease progression. Blue peaks denote copy number gain and red peaks denote copy number loss.

The most frequent abnormality observed at relapse was copy number gain of 1q21.2-1q23.3, which is within a region identified by GISTIC2.0. The region was gained in 9/27 (33.3%) of relapse cases and 7/58 (12.1%) in the non-progression group (p=0.034, Fisher exact test). 7q gain was also slightly more common in the poor outcome group with 3/27 cases (11.1%) compared to 2/58 (3.5%) in the group without events (p=0.321, Fisher exact test), although this was not statistically significant. This is noteworthy as cytogenetic analysis of the FAB/LMB96 trial cohort highlighted 7q gain as prognostic but it was not prognostic in the UK B-NHL cohort in this analysis (Poirel *et al.*, 2009). Gain of chromosome 12 was only seen in cases without disease progression, present in 13.8% of this group. The disparity in incidence was not significant (p=0.051, Fisher exact test) but may highlight an important role for genes on chromosome 12 in disease progression and sensitivity to treatment.

Chapter 4 highlighted three significant prognostic markers in paediatric B-NHL; 17p CNN-LOH, 17q CNN-LOH and gain of 3q29 and all three aberrations were observed in the eBL cohort. 17p CNN-LOH was significantly associated with relapse in the sporadic cohort while it was present in only 3.7% of relapse cases (1/27) in eBL and 5.2% of cases without disease progression (3/58, p=1, Fisher exact test). It is noteworthy that 17p CNN-LOH was not associated with disease progression in eBL when it is a strong prognostic marker in a cohort consisting of primarily the sporadic subtype. This highlights the important of identifying subtype-specific prognostic markers and may refute the idea that prognostic markers in one BL subtype are applicable to others, as suggested by Panea *et al.* (2019).

Conversely to 17p aberration, 17q CNN-LOH was more common in the relapse group of the eBL cohort, although this association with relapse was not significant (7/27, 25.9% vs 7/58, 12.1% p=0.13, Fisher exact test). 17q CNN-LOH in the eBL cohort was most frequently affecting 17q23.1-17q25.3 and the telomere, overlapping the 17q22-17q24.1 region seen in the sBL cohort and including the *GNA13* gene, characterised in Chapter 4.

3q29 gain was common in eBL as well as being a marker of poor outcome in sBL (HR 3.8 95% CI 1.1-13, p=0.041, Cox proportional hazard method). In contrast, gain

at 3q29 was most common in the patients without disease progression (9/58, 15.5% vs 1/27, 3.7%, *p*=0.16, Fisher exact test). Only one relapse case had gain at 3q29 itself, but there was a peak in five relapse cases with copy number gain involving *BCL6* at 3q27.3-3q28. The hazard ratio of 3q29 gain in eBL was 0.26 (95% CI 0.035-1.9, *p*=0.18, Cox proportional hazard method).

6.3.5 Univariate survival analysis of genomic and clinical factors in eBL

Univariate survival analysis was performed by Cox proportional hazards method on genomic markers identified in eBL. For the endemic cohort two survival metrics were utilised: overall survival (OS) and risk of relapse (RR). RR follow up was defined as the time from diagnosis to either last follow up or relapse event. RR was used instead of TTP to investigate genomic associations with relapse specifically as only patients who completed frontline therapy were included as part of RR analysis. Refractory disease was excluded from analysis. Survival estimates for the cohort at 12 months was 61.3% (CI 95% 50.9-73.9) for RR analysis and 64.3% (CI 95% 54.4-76.0) for OS (Figure 6.8A-B). There were 85 patients with sufficient data for RR analysis and 86 with data to perform OS analysis.

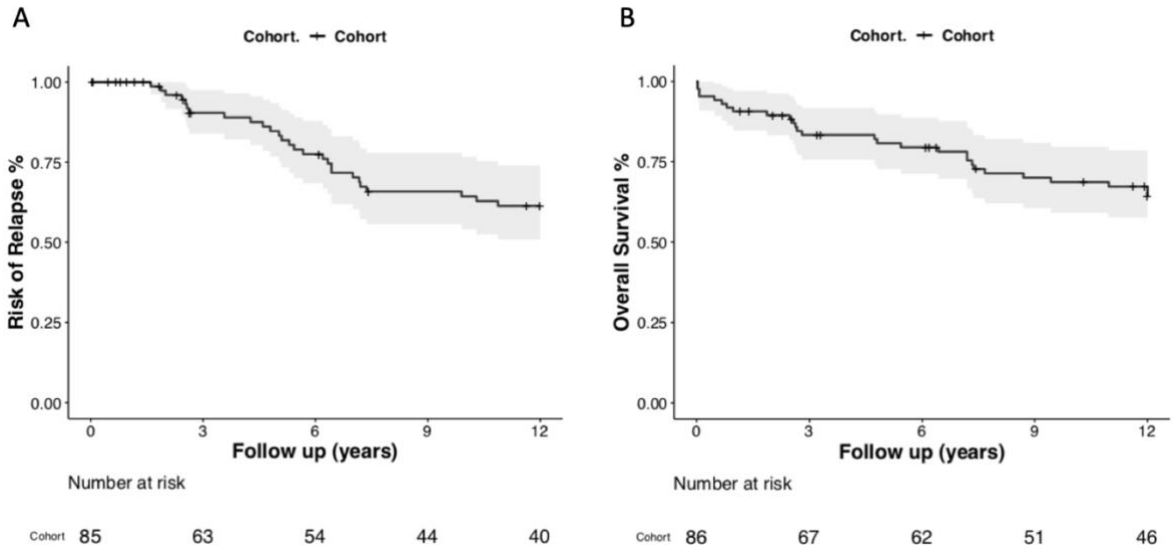


Figure 6.8 - Kaplan Meier plots showing (A) Risk of Relapse Survival and (B) Overall Survival for the eBL cohort.

Focal regions identified by GISTIC2.0 in the eBL cohort were included as univariate factors as well as patient gender, the three prognostic markers identified in chapter four (17p CNN-LOH, 17q CNN-LOH and 3q29 gain) and gain of two genes of interest; *MCL1* and *FOXO1*. *MCL1* is a putative gene of interest in the gained region

of chromosome arm 1q and overexpression in B-NHL cell lines has been shown to exert resistance to multiple chemotherapeutic agents (Vrana *et al.*, 2002). Amplification of *MCL1* was a recurrent abnormality reported by Panea *et al*. (2019) associated with endemic BL. *FOXO1* gain was commonly observed but segments of copy number gain covering the region were considered to be the result of noise and not considered a somatic event. Previous work in the lab has shown an important role for *FOXO1* in eBL so this aberration was included in analysis (Zhou *et al.*, 2019). As expected, prognostic markers identified in Chapter 4 3q29 gain, 17p CNN-LOH and 17q CNN-LOH were all significant at univariate analysis in this sporadic BL cohort (Table 6.4).

Factor	HR (95% CI for HR)	*p*.value
17q CNN-LOH	4.8 (1.5-15.0)	0.007
17p CNN-LOH	3.6 (1.1-11.0)	0.029
3q29 Gain	3.8 (1.1-13.0)	0.041

Table 6.4 - Genomic abnormalities associated with prognosis in chapter four were significantly associated with a poorer time to progression by univariate survival analysis in the sporadic BL cohort.

Three factors were identified to be significantly associated with higher risk of relapse by univariate survival analysis with a *p* value $\leq$ 0.1; 17q11.2 gain, patient gender and *MCL1* gain. Genomic abnormalities with prognostic value identified by univariate Cox Proportional Hazard modelling are summarised in Table 6.5. Prognostic markers where p<0.1 in univariate survival analysis were taken forward for multivariate modelling.

Factor	HR (95% CI for HR)	*p* value
17q11.2 Gain	3.1 (1.1-9)	0.038
Gender (F)	3.4 (1.2-9.5)	0.088
MCL1 Gain	2 (0.88-4.6)	0.097

Table 6.5 - Significant factors by univariate survival analysis in the eBL cohort.

Patient gender was shown to be associated with relapse in the eBL cohort, with female patients having the highest risk of relapse (Figure 6.9A). The gender was not prognostic in itself, but the p value was below the threshold (detailed in Chapter 2.4.16) to be included in multivariate modelling by forward selection.

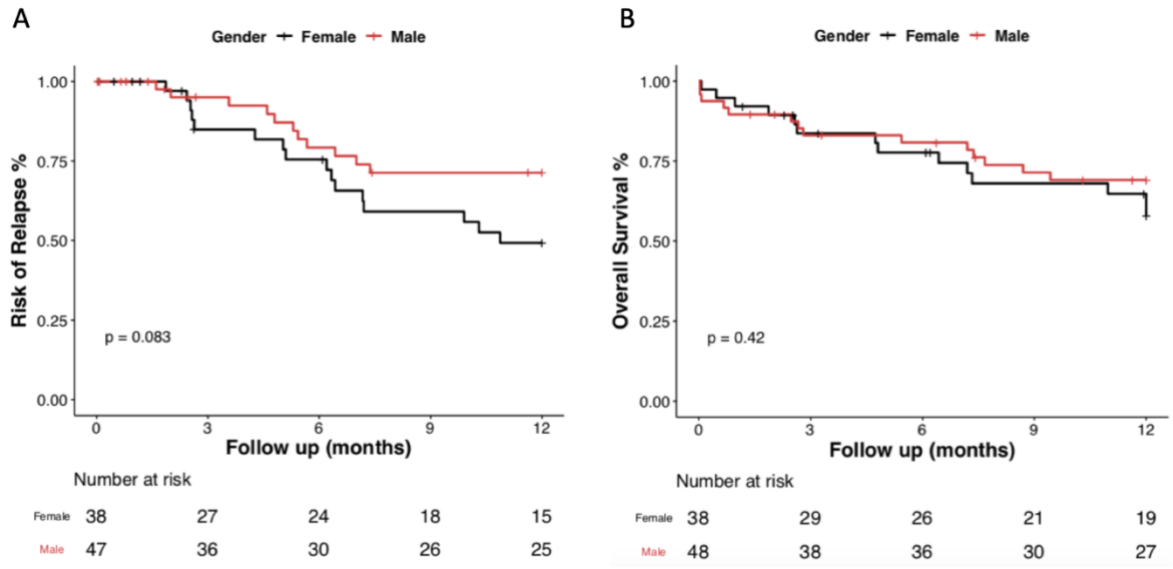


Figure 6.9 - Kaplan Meier plots showing (A) Risk of Relapse Survival (p=0.083, Log rank test) and (B) Overall Survival (p=0.420, Log rank test) for gender.

Of 105 patients in the cohort, 58 patients were male and 47 were female and the hazard ratio for female patients was 3.4 (CI 95% 1.2-9.5, *p*=0.088, Cox proportional hazard method). This suggests that patient gender is not a significant prognostic marker but may improve a multivariate model. The separation of Kaplan Meier curves suggests that any prognostic value of patient gender is small and the correlation was not seen in OS analysis (Figure 6.9B).

17q11.2 gain was observed infrequently in the cohort (6/105, 5.7%) but was associated with a significantly higher risk of relapse with a HR 3.1 (CI 95% 1.1-9, *p*=0.038, Cox proportional hazard method). The region involved 52 genes and was 2.1Mb in size. Of the six patients with copy number gain of this region, four relapsed and three died (Figure 6.10A-B).

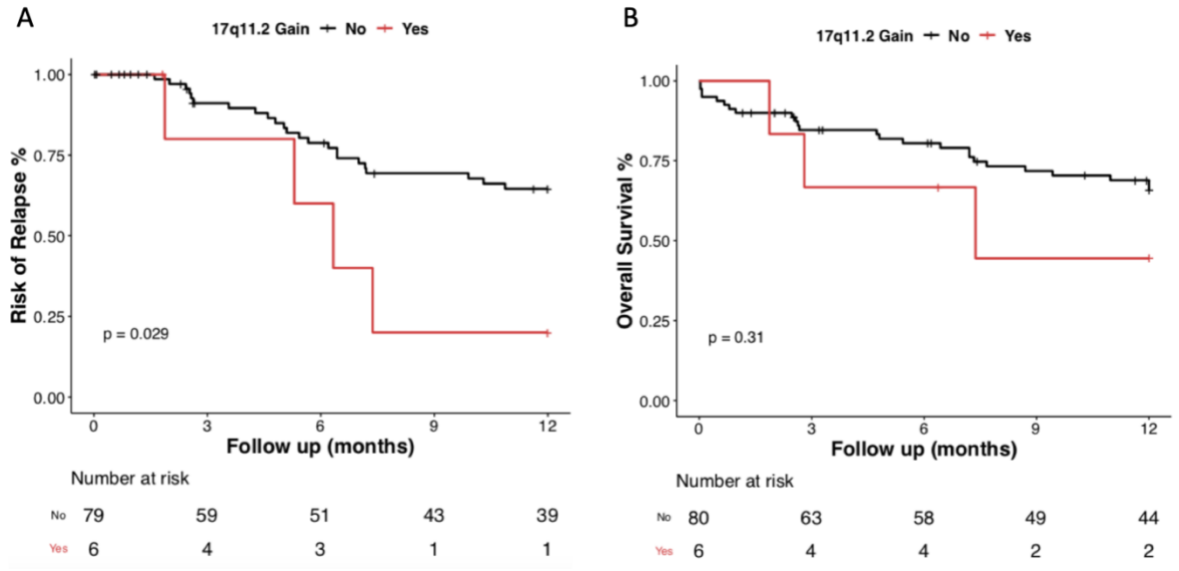


Figure 6.10 - Kaplan Meier plots showing (A) Risk of Relapse Survival (p=0.029, Log rank test) and (B) Overall Survival (p=0.310, Log rank test) for 17q11.2 gain.

The region of gain at 17q11.2 identified by GISTIC2.0 and significant by univariate survival analysis covered 52 genes (Figure 6.11). Gene ontology analysis using Kegg Pathway Mapper (Kanehisa *et al.*, 2019) highlighted the *MIR451A* tumour suppressor gene, which has been implicated in drug resistance in breast cancer. The relevance of copy number gain of this gene in eBL is unclear. The *TRAF4* gene was also gained in these samples and is known to drive metastasis in breast cancer through TGF-β signalling (Zhang *et al.*, 2013).

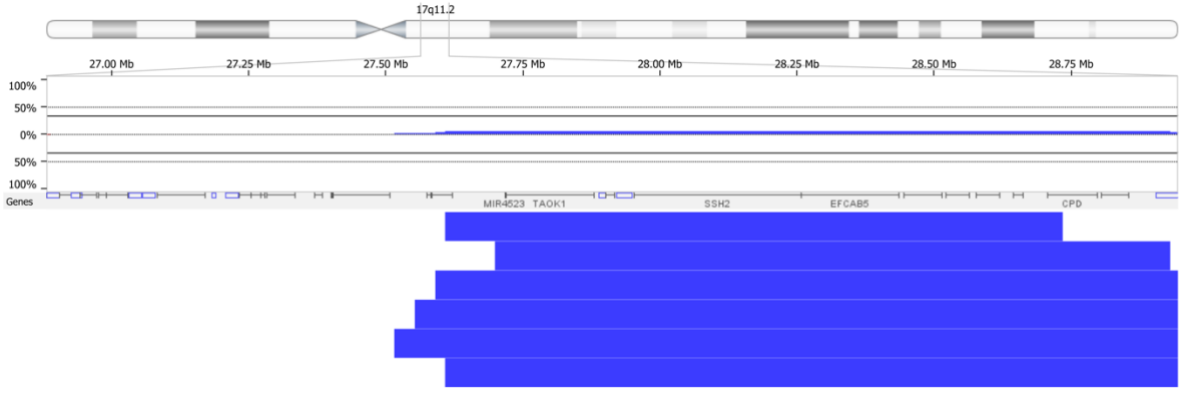


Figure 6.11 – A copy number plot showing copy number gain of 17q11.2 in eBL. Blue denotes copy number gain and red denotes copy number loss.

MCL1 gain on cytoband 1q21.3 was recurrently seen in both sBL and eBL as part of larger 1q copy number gains and was present in the second most significant region of amplification identified by GISTIC2.0. Gain of the gene itself was associated with a

higher risk of relapse in the eBL cohort with a HR of 2.0 (CI 95% 0.88-4.6, *p*=0.097, Cox proportional hazard method), but similarly to gender, this may only be significant in the context of a multivariate model (Figure 6.12A-B).

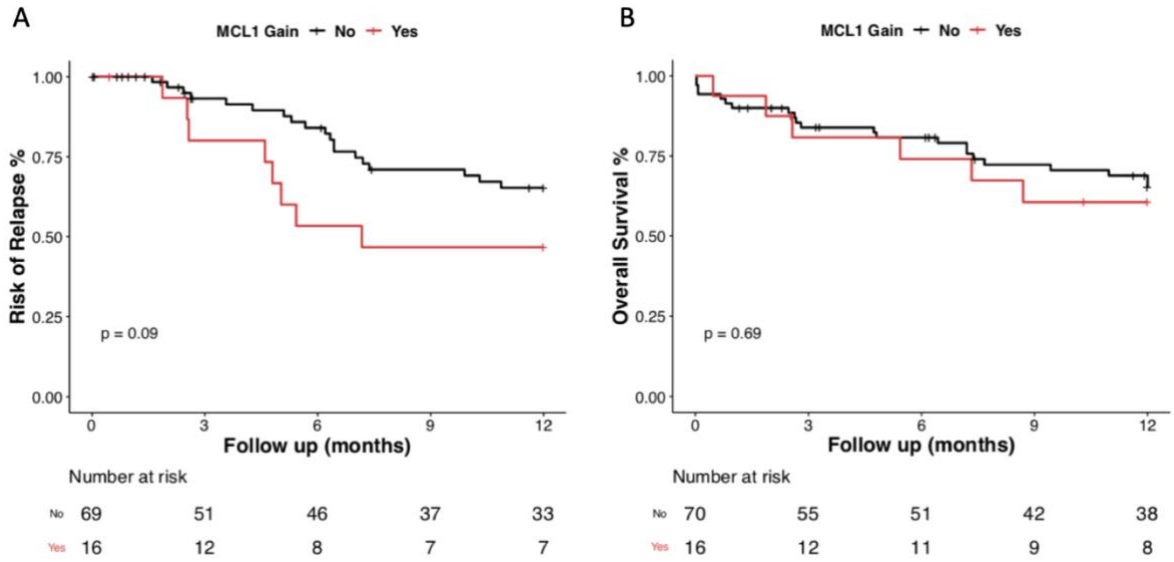


Figure 6.12 – Kaplan Meier plots showing (A) Risk of Relapse Survival. (p=0.090, Log rank test) and (B) Overall Survival (p=0.690, Log rank test) for MCL1 gain.

The *MCL1* gene is well-studied in B-NHL and has been shown to be important in sensitivity and resistance to therapy (Vrana *et al.*, 2002). Previous studies have shown that loss of *MCL1* in B-NHL cell lines results in sensitivity to BCL2 inhibitor Venetoclax (Phillips *et al.*, 2015). *MCL1* was gained in 21.0% of the eBL. Patterns of gain ranged from small focal events to gain of the whole chromosome arm (Figure 6.13).

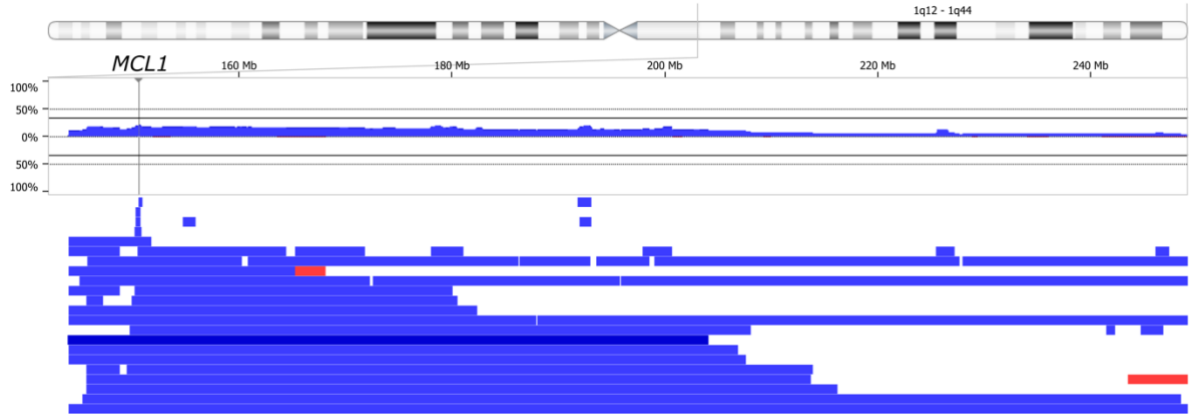


Figure 6.13 - A copy number plot showing copy number gains involving the MCL1 gene on 1q in eBL. Blue denotes copy number gain and red denotes copy number loss.

Univariate analysis was repeated on a subset of the eBL cohort treated more uniformly on the NL3 protocol with doxorubicin to confirm whether these markers were still prognostic in a cohort treated using current treatment protocols (Appendix G). In 71/76 (93.4%) patients treated with the NL3 protocol, RR analysis identified gender as the only significant marker of higher risk of relapse with the HR of female patients as 3.4 (CI 95% 1.2-9.5, p=0.021, Cox proportional hazard method). In contrast, *MCL1* gain and 17q11.2 gain were not associated with RR in patients uniformly treated on NL3 protocols (p=0.67 and 0.14 respectively, Fisher exact test).

6.3.6 Multivariate survival analysis

Multivariate analysis was performed on the whole eBL cohort regardless of treatment protocols using factors with a p value of 0.1 or less by univariate survival analysis of eBL patients; gender, 17q11.2 gain and *MCL1* gain (Table 6.5). These factors were included as they were significant at p<0.1 by univariate analysis. Model fit 1 (gender and 17q11.2 gain) was compared against model fit 2 (17q11.2 gain and *MCL1* gain). Model 2 had a more significant likelihood ratio test p value at p=0.01 compared to p=0.03, meaning that it was the better model. Adding gender to the model to create model 3 improved the likelihood ratio p value to p=0.007 (Table 6.6).

Covariate	Hazard ratio (95% CI)	*p* value
17q11.2 Gain	1.24 (1.15-10.40)	0.027
MCL1 Gain	1.00 (1.15-6.51)	0.023
Gender (F)	0.76 (0.97-4.72)	0.058

Table 6.6 - Cox proportional hazard multivariate model 4 in eBL patients.

Comparing the models with using the likelihood ratio test also suggested that model 3 may have been a better than model 2, although the p value was not significant (p=0.052).

6.3.7 TP53 abnormalities in eBL

The *TP53* status of BL was determined to be the most important genomic factor in predicting and driving relapse in paediatric B-NHL and specifically sporadic BL in Chapter 5. The effect of *TP53* abnormalities on outcome in the endemic subtype is less well understood. A study by Grande *et al. (2019)* showed that *TP53* mutations

were less frequent in EBV positive BL cases compared to EBV negative BL, but no study has investigated *TP53* aberrations in the context of patient outcome or relapse.

17p deletion involving *TP53* was observed at similar incidences in both subtypes (12/79, 15.2% sBL and 16/105, 15.2% eBL, p=1, Fisher exact test). In contrast, 17p CNN-LOH affected 10.1% (8/79) of sBL patient samples and 3.8% (4/105) of eBL (p=0.130, Fisher exact test). In the sporadic cohort 17p CNN-LOH was associated with an inferior time to progression (HR 3.6 95% CI 1.1-11.0, p=0.029, Cox proportional hazard method). The hazard ratio in the eBL cohort was not calculable, as there were too few cases in the relapse group.

Work performed in parallel to this project by Dr Peixun Zhou using Sanger sequencing had characterised the *TP53* gene mutations in the cohort for 101/105 cases (96.2%). In total, 40 mutations were identified in 37/101 diagnostic samples screened (36.6%) and 3/37 samples had a second mutation. By Sanger sequencing it was not possible to determine whether two mutations in a sample were on the same allele. However, the assumption was made that this was the case given the evidence observed in the sporadic cohort in Chapter 5, where two cases were shown to have compound heterozygous mutations in *TP53*. The *TP53* mutation status was known for 70/79 sBL cases, with 42/70 (60.0%) harbouring at least one mutation, and 7/42 having two mutations, totalling 49 mutations identified. Of these, two were R213R silent mutations and the remainder were missense mutation. The locations of mutation in *TP53* in both cohorts are summarised in Figure 6.14 and Appendix H.

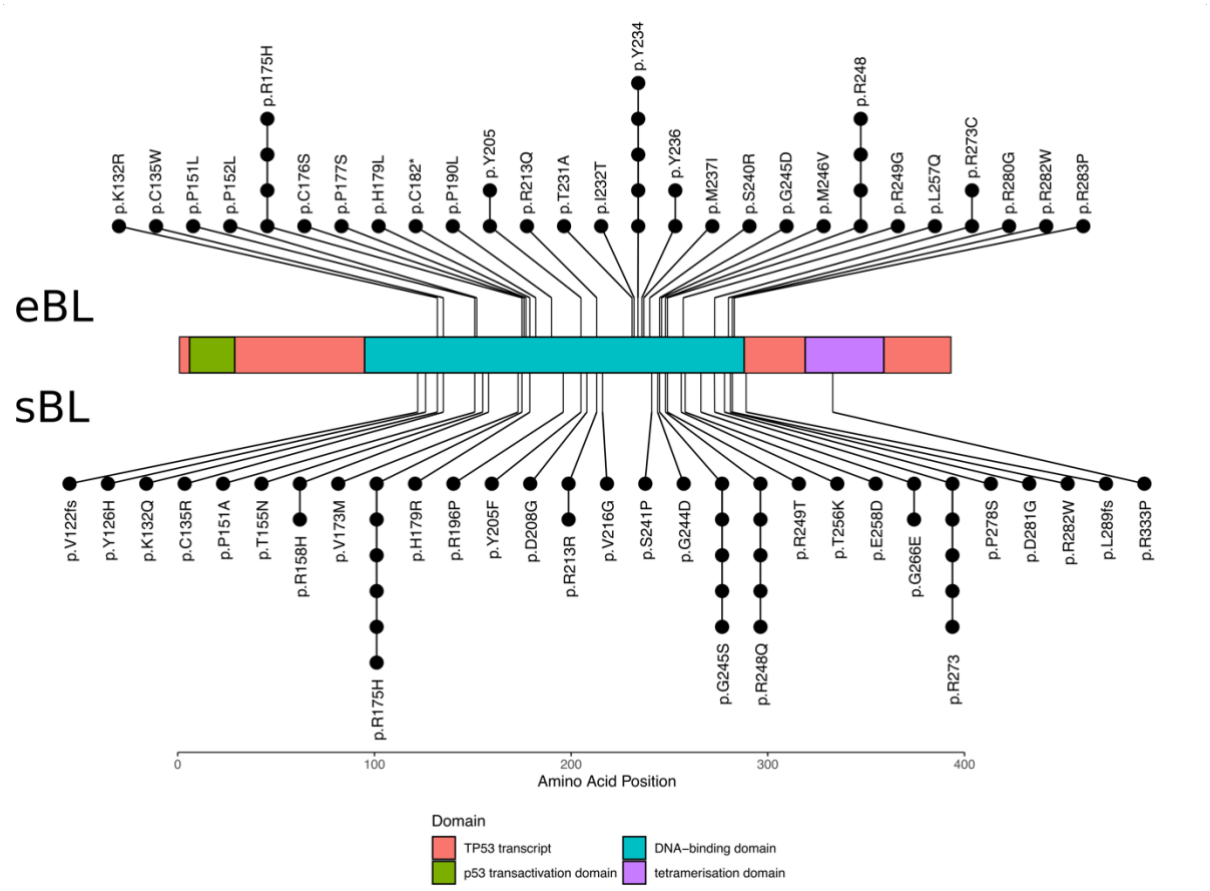


Figure 6.14 - A lollipop plot showing TP53 mutations identified in eBL and sBL.

The incidence of *TP53* mutation was significantly higher in the sporadic cohort (42/70, 60.0% vs 37/101, 36.3%, p=0.003, Fisher exact test) which was consistent with the findings of both the Grande *et al.* (2019) and Panea *et al.* (2019) studies. In sBL the most common mutations were R175H, G245S, R248Q and variants at R273.While all four were observed in the eBL cohort, G245 mutation was only seen in one case, with a different amino acid change to those seen in the sporadic cohort. The most frequently mutated residue in the eBL cohort was Y234, which was not seen in any sporadic patients. Y234 mutations were found in the IARC *TP53* database and reported to give rise to non-functional p53 protein (Petitjean *et al.*, 2007). Other mutations in surrounding residues were seen infrequently between amino acids 231 and 240 in the eBL cohort, but the area did not harbour mutations in the sporadic cohort.

Integrating this data with the array data to take 17p deletion and CNN-LOH into account allowed for investigation into the clinical relevance of biallelic and monoallelic *TP53* aberrations in eBL compared to sBL. In total, 47/101 (46.5%) eBL

patients had *TP53* aberrations, with 35 (34.7%) having only one monoallelic abnormality and 12 (11.9%) having biallelic aberrations. 54/101 (53.5%) cases had no detectable *TP53* aberrations. Of those with monoallelic aberrations 25/35 (71.4%) had a single mutation and 10/35 (28.6%) had deletion without mutation. Of the biallelic cases, 4/12 (33.3%) had a single mutation and with CNN-LOH, 6/12 (50.0%) had hemizygous mutations alongside 17p deletion. A further two biallelic cases harboured two mutations which were considered to be compound heterozygous. In the sporadic cohort 44/70 (62.9%) had at least one *TP53* aberration, with 23/44 (52.3%) having only one monoallelic aberration and 21/44 (47.7%) harbouring biallelic abnormalities of the *TP53* gene. *TP53* abnormal cases were significantly less frequent in eBL compared to sBL (47/101, 46.5% vs 44/70, 62.9% p=0.043, Fisher exact test).

Biallelic aberration of the *TP53* gene was relatively infrequent in the eBL cohort compared to the sBL group (12/101, 11.9% vs 21/70, 30%; p=0.005, Fisher exact test). In contrast, 23/70 (32.9%) sporadic patients had monoallelic abnormalities which was consistent between cohorts (35/101, 34.7% vs 23/70, 32.9%, p=1, Fisher exact test). The presence of any *TP53* aberration was not associated with risk of relapse or overall survival in eBL by Cox Proportional Hazards modelling (RR HR 1.1 95% CI 0.5-2.3, p=0.89, OS HR 1.1 95% CI 0.5-2.3, p=0.81, Cox proportional hazard method) (Figure 6.15A-B).

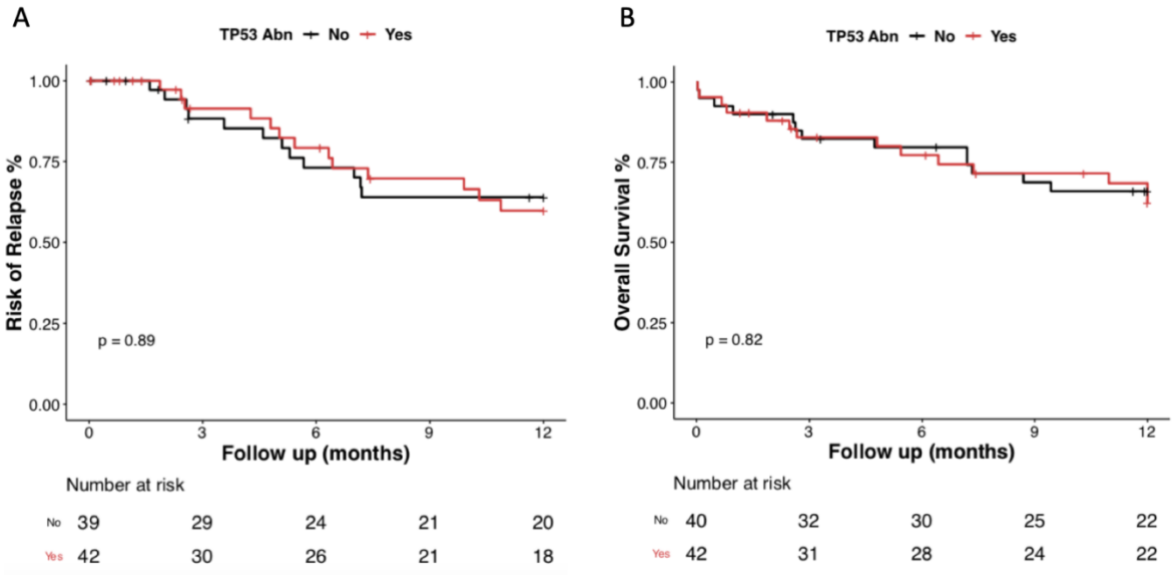


Figure 6.15 - Kaplan Meier plots showing the (A) Risk of Relapse (p=0.89, Log rank test) and (B) Overall Survival (p=0.82, Log rank test) analysis for eBL patients with TP53 abnormalities.

Biallelic or monoallelic status were also not associated with an inferior risk of relapse when compared against patients without any *TP53* aberrations (Biallelic RR HR 1.1 95% CI 0.3-4.0, *p*=0.84, Monoallelic RR HR 1.0 95% CI 0.5-2.4, *p*=0.93, Cox proportional hazard method). Despite this, it was shown by OS analysis that *TP53* biallelic cases eBL do worse than both monoallelic and *TP53* normal subgroups (HR 2.3 (95% CI 0.9-6.0), *p*=0.098, Cox proportional hazard method) (Figure 6.16).

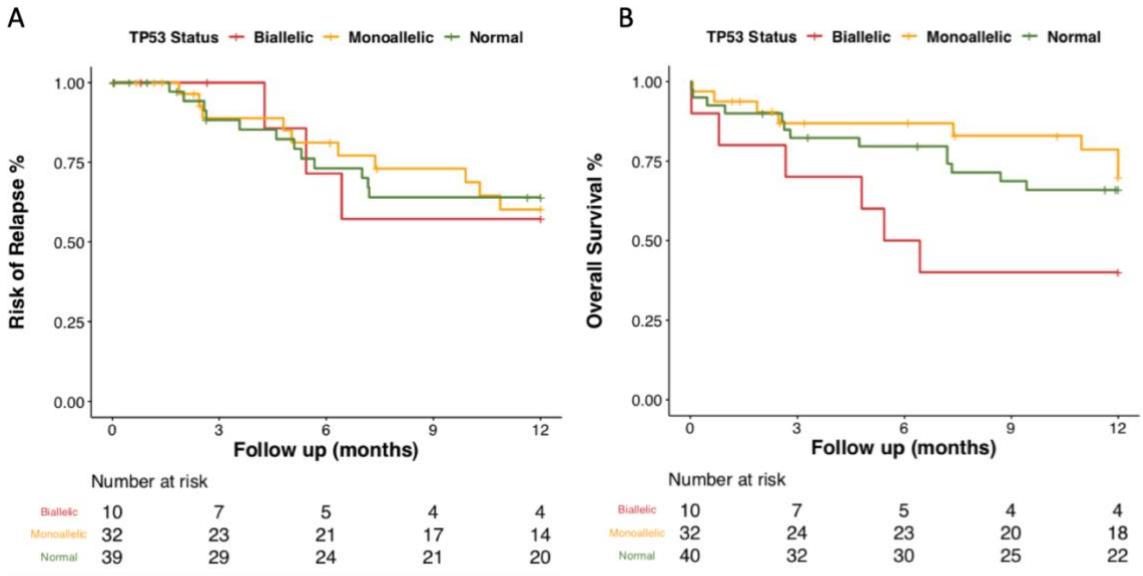


Figure 6.16 - Kaplan Meier plots showing the (A) Risk of Relapse (Biallelic vs Normal p=0.842, Biallelic vs Monoallelic p=0.875, Monoallelic vs Normal p=0.932, Log rank test) and (B) Overall Survival (Biallelic vs Normal p=0.089, Biallelic vs Monoallelic p=0.043, Monoallelic vs Normal p=0.590, Log rank test) analysis for eBL patients with monoallelic, biallelic and normal TP53 status.

The association was not significant when comparing biallelic cases against *TP53* normal cases, but it was significant when comparing the biallelic patients against the rest of the cohort, including monoallelic and normal cases (HR 2.5 95% CI 1.0-6.2, *p*=0.048, Cox proportional hazard method). Comparing the impact of *TP53* abnormalities on outcome in the endemic and sporadic subgroups showed that the presence of *TP53* biallelic aberrations in eBL does not predict relapse disease in the endemic setting but that regardless of relapse status, these patients were more likely to die of disease. This is in contrast to sBL where biallelic *TP53* status was shown to be a potential prognostic marker for disease progression but was not associated with outcome (Figure 6.17A-B).

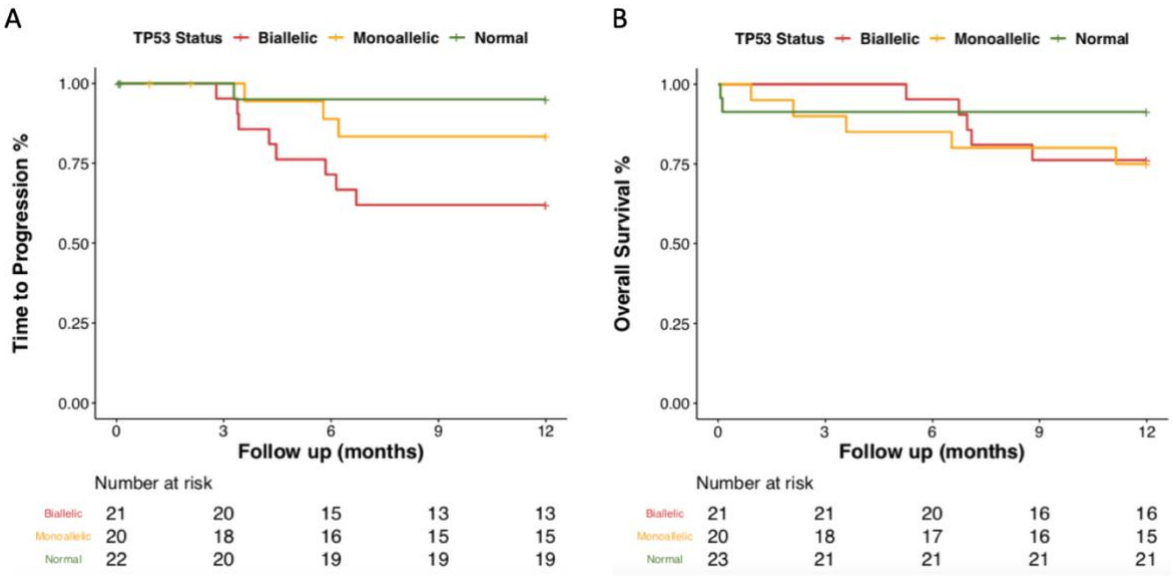


Figure 6.17 - Kaplan Meier plots showing the (A) Time to progression (Biallelic vs Normal p=0.014, Biallelic vs Monoallelic p=0.133, Monoallelic vs Normal p=0.266, Log rank test) and (B) Overall Survival (Biallelic vs Normal p=0.208, Biallelic vs Monoallelic p=0.843, Monoallelic vs Normal p=0.183, Log rank test) analysis for sporadic BL patients with monoallelic, biallelic and normal TP53 status.

6.3.8 Copy Number Aberrations associated with TP53 status in eBL

Copy number profiles for eBL patients with *TP53* abnormalities were compared with the profiles of patients with normal *TP53* (Figure 6.18) and statistical analysis performed by Fisher's exact test.

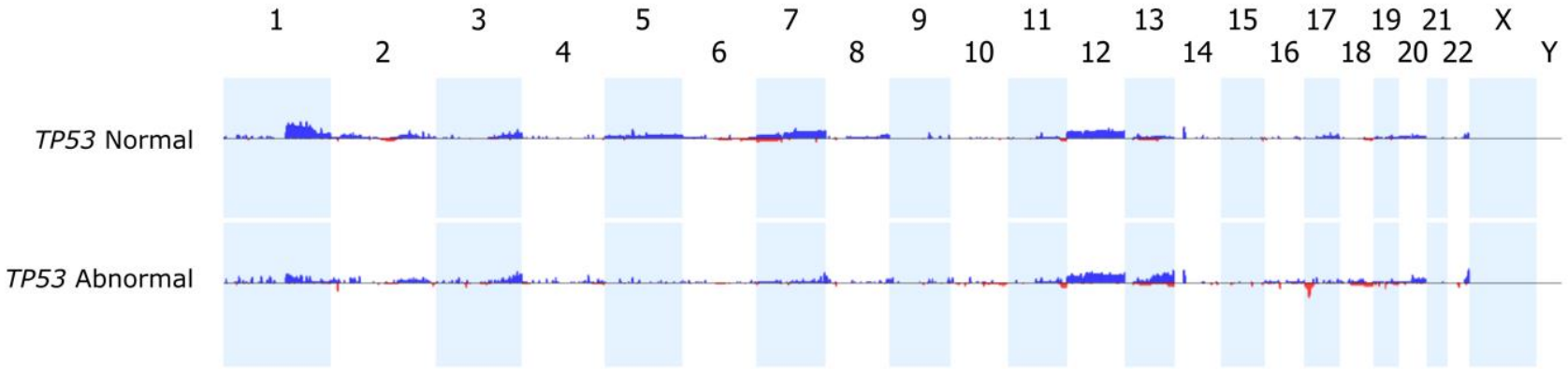


Figure 6.18 - Copy number aggregate plot showing the copy number profiles of eBL patients both with and without TP53 abnormalities. Blue peaks denote copy number gain and red peaks denote copy number loss.

Two regions identified by GISTIC2.0 analysis in the eBL cohort were significantly associated with *TP53* status. The 11.7Mb region spanning 13q31.1 to 13q32.2 was significantly associated with *TP53* abnormality (8/47, 17.0% vs 2/54, 3.7%; p=0.04, Fisher exact test). This correlates with the finding that *MIR17HG* amplification and $13q^{plex}$ pattern were associated with *TP53* aberration in sBL in chapter 5. Additionally, focal 1q31.2-1q31.3 gain was recurrent in the *TP53* normal group. Larger 1q gains were also more common in the *TP53* normal group (12/54, 22.2% vs 4/47, 8.5%; p=0.1, Fisher exact test). Gain of 7q most commonly occurred in the *TP53* normal group (6/54, 11.1% vs 1/47, 2.1%; p=0.12, Fisher exact test). While these abnormalities were not significantly associated with *TP53* status, there was a clear association between both 1q and 7q gain with *TP53* in eBL which should be confirmed in a large cohort of patient samples.

6.4 Discussion

Analysis performed in this chapter determined that *TP53* aberrations were markedly less common in endemic BL when compared to sporadic BL, which supports the hypothesis that *TP53* abnormality plays a diminished role in endemic BL. This is further supported by the lack of association between *TP53* status and relapse in the eBL cohort, in contrast to sporadic BL. EBV infection in BL is known to inhibit MYC-driven pro-apoptotic signals, including p53-mediated apoptosis. EBNA3C nuclear antigen coded for by the EBV genome acts to inhibit p53 function, including p53-mediated apoptosis via interaction with and stabilisation of ING4, ING5 and MDM2 protein (Saha *et al.*, 2009). Considering this in the context of the findings of our study, it is possible that *TP53* aberrations are largely redundant in EBV positive BL, with p53 function abrogated through viral mediated mechanisms (Fitzsimmons and Kelly, 2017). It is remarkable that patients with biallelic *TP53* abnormalities had a significantly worse overall survival, suggesting that even in the presence of EBV infection *TP53* plays an important role in chemotherapy resistance. The poor outcome for these patients is to be expected, as a course of intensive chemotherapy in the form of FAB/LMB96 protocols with extensive supportive care is insufficient to treat this subgroup of *TP53* biallelic cancers, as demonstrated by analysis in Chapter 5. Whole genome and transcriptomic characterisation of these patients from both patients would allow for a better understanding of the prognostic role of *TP53* and determine whether *TP53* biallelic subgroup of eBL with a poorer overall survival resemble the cases seen in the sporadic cohort that undergo disease progression.

This study investigated the genomic profiles of 105 eBL patients with copy number array data. Recurrent abnormalities involving 1q, 7q and 13q were consistent with what was previously identified in sporadic B-NHL cases in chapter 4, and specifically sBL in this chapter. Copy number analysis also highlighted regions which were associated more strongly with a subtype. For example, 6q loss was recurrent in sBL but almost entirely absent in the eBL cohort. Conversely, gain of chromosome 12 was common in the eBL cohort but seen in only one sBL patient. As would be expected, the genomic aberrations linked to poor outcome in eBL differed to those identified in sBL. None of the three prognostic markers identified in sporadic BL or the larger sporadic B-NHL cohort were shown to be prognostic in the endemic cohort.

Underlying biological differences associated with EBV infection may also explain the disparity of prognostic markers between disease subtypes. There is evidence that, in BL, EBV infection promotes genomic instability, including the occurrence of complex chromosomal rearrangements similar to those observed in cases with *TP53* aberrations in the sporadic cohort in Chapter 5 (Kamranvar *et al.*, 2007). Genomic complexity in the sporadic B-NHL cohort (Chapter 5) was not itself associated with poorer outcome, so even if EBV and *TP53* aberration play complementary roles in driving genomic instability, this may not fully explain the difference in outcome between cohorts.

Treatment of endemic Burkitt lymphoma is a significant clinical challenge and an important focus for the field of B-NHL research. A combination of a lack of large cohort genomic studies into eBL, a reduced capacity for high-intensity chemotherapy and unique socioeconomic variables creates a complex clinical problem. Survival rates have been increasing steadily over recent years, particularly after the introduction of doxorubicin to treatment protocols in Malawi for stage III and IV disease (Molyneux *et al.*, 2017). Despite this, rates of survival have fallen short of the results seen in sporadic BL in the UK. A recent study was undertaken to determine the feasibility of eBL being treated with “standard chemotherapy” in a resource constrained setting, with greater efforts made to provide supportive care and medication to counteract side effects of the chemotherapy, reduce malnutrition in patients and supported patients in transportation to reduce abscondment (McGoldrick *et al.*, 2019). While supportive care was successfully administered to paediatric patients receiving treatment with intense chemotherapy, 4-year OS for the cohort remained low at 44%. Unless paediatric eBL can be successfully treated using intense FAB/LMB-based protocols then there will be a clinical need for prognostic markers specific to eBL.

3q29 was the most significantly gained region identified by GISTIC2.0 analysis of the genome in endemic BL in this analysis. The aberration also had no impact on prognosis, despite being identified as a strong predictor of inferior time to progression and lower overall survival in sporadic B-NHL in chapter 4. The region of gain seen in each subtype were comparable in size and both involved the 3q telomere. The presence of the aberration in both cohorts suggests that 3q29 gain is important in a

pathway or process common to BL subtypes. Scholtysik *et al*. (2010) suggested that copy number gains involving the 3q telomere in their cohort were likely driving *BCL6* expression which is located at 3q27.3. However the majority of 3q29-gained cases in both cohorts associated with this analysis, as well as in the analysis in Chapter 4, did not involve copy number gain of this gene. Elucidation of a driver gene within this region will pave the way for better understanding of universal BL biology and the nature of treatment resistance in different subtypes. 17q CNN-LOH was also recurrent in eBL but shown not to be associated with progression or death in the endemic setting.

A major drawback of investigating the role of *TP53* and other genomic markers in disease progression and patient outcome is the analysis of diagnostic samples alone. Tumours evolve with time and treatment plays a key role in providing selection pressures for development of the tumour. *TP53* status at diagnosis may not be maintained at the point of disease progression. It is entirely feasible that endemic BL cases with no detectable *TP53* mutation, loss or CNN-LOH may acquire additional aberrations involving the gene which change the clonal makeup of the tumour, spurring therapy resistance. In Chapter 7 a panel of B-NHL patients with paired diagnostic and relapse material is analysed to further investigate this question. Four of these patients were sporadic BL cases from the *TP53* biallelic high-risk subgroup investigated in this chapter as well as Chapter 5. Copy number analysis of these patients as their chemotherapy-resistant disease progresses will increase our understanding of prognostic markers in paediatric B-NHL.

Chapter 7: Clonal Evolution of Progressive Disease

7.1 Introduction

The genetic landscape of disease progression in paediatric B-NHL is poorly understood. The lack of understanding and research into the mechanisms of disease progression in paediatric B-NHL is primarily due to a lack of usable sequential biopsies available for research. Even when a diagnostic biopsy is banked from a case who goes on to relapse it is still often uncommon for the sequential biopsy or biopsies to be banked. The clinical team generally take minimal tissue at biopsies of progressive disease leaving very little to be banked for research purposes. In this study material was acquired for 11 B-NHL patients at diagnosis and relapse.

Studies into the clonal evolution of DLBCL by VDJ sequencing identified two distinct patterns of relapse (Jiang *et al.*, 2014). A pattern of early-divergent or “branching” evolution involving two distinct clones at diagnosis and relapse with a common ancestor arising early in disease development according to phylogenetic analysis of somatic hypermutation (SHM) data (Figure 7.1A), and a second pattern of late divergent “linear” evolution, where the diagnostic and relapse major clones closely resemble each other (Figure 7.1B).

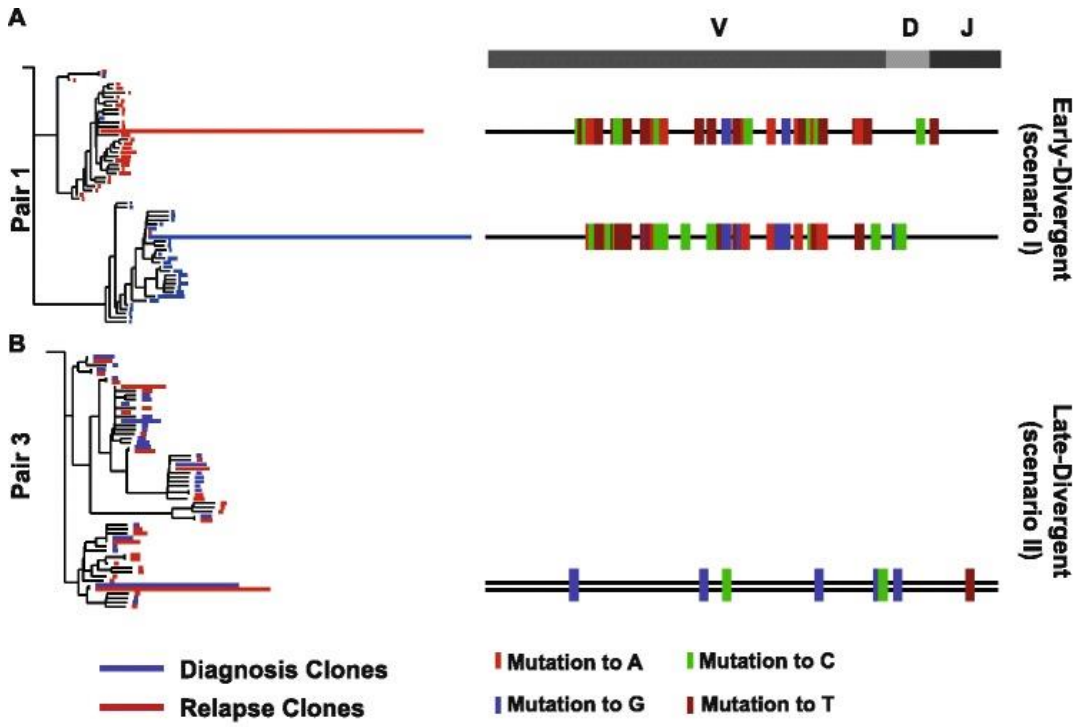


Figure 7.1 – Visual representation of two patterns of disease progression and tumour evolution in DLBCL. (A) The Early-divergent scenario involves two clones that have a shared ancestor but do not strongly resemble each other. (B) The Late-divergent scenario shows a major clone at relapse which shares many features with the diagnostic major clone. From Jiang et al. *(2014).*

Comparison of the mutational distance between diagnostic and relapse clones revealed that tumours with the early-divergent model of progression had significantly higher rates of SHM at relapse compared to diagnosis. There was no association between DLBCL subtype and pattern of evolution, although GCB-like DLBCL patients did relapse significantly later than ABC-like DLBCL patients (Jiang *et al.*, 2014). A later study of 20 DLBCL diagnostic and relapse pairs successfully determined the evolutionary pattern of 18/20 pairs from copy number array data (Juskevicius *et al.*, 2016). Characterisation of genetic determinants of evolutionary pattern may offer new targets for therapy.

The most significant study of disease progression and evolution in sequential BL samples investigated 6 paediatric and 1 adult cases with karyotyping data (Aukema *et al.*, 2015). The major findings were that samples taken at the point of disease progression had significantly more complex karyotypes and that BL cases progress in a linear/late-branching pattern of evolution. A later study of relapsed/refractory DLBCL pairs postulated that late-divergence is a strategy of tumour survival in response to the selection pressure of therapy (Juskevicius *et al.*, 2017) and as such

is not a pre-existing feature of the tumour but is acquired rapidly during treatment, regardless of B-NHL subtype. This would preclude the relevance of prognostic markers, as by this definition the genomic driver of chemotherapy resistance would not be present at diagnosis. The prognostic markers identified as part of Chapters 4 and 5 were significant in a large cohort of patients, particularly in BL cases. This is significant, considering that Aukema *et al.* (2015) report that over 80% of BL relapse/refractory cases exhibited a late-branching/linear pattern of evolution. The fact that *TP53* aberration, as well as 3q29 copy number gain and 17q CNN-LOH were significant, independent prognostic markers is incompatible with the concept of late acquisition of genomic aberrations driving treatment resistance in BL. One possible explanation for this is that prognostic markers identified at diagnosis in our cohort represent enabling events that predispose the tumour to further evolution through genomic instability, evidence for which is provided in Chapter 5.

In Chapter 4, the largest clinically annotated cohort of paediatric B-NHL patients with copy number data was analysed to identify genomic markers of prognosis. Besides *MYC* rearrangements, all copy number abnormalities that were detected were present in a relatively small numbers of cases (<30%) and only three abnormalities were found to be prognostic in the sporadic cohort. Conversely, in Chapter 5 we showed that aberrations of the *TP53* gene, including mutations, were present in over 50% of cases, and were significantly associated with poor outcome. This suggests that the vast majority of copy number aberrations in paediatric B-NHL may be passenger abnormalities that do not play a vital role in maintenance of the malignancy or drive disease progression in poor outcome cases.

In Chapter 6 it was reported that in endemic BL *TP53* aberrations had a less significant impact in outcome. While *TP53* biallelic aberrations were significantly associated with poor outcome, there was no association between abnormalities involving *TP53* and relapse. The most likely explanation for this is the different treatment received by each cohort.

It is important to investigate the genomic profile of tumours throughout cancer development and progression in order to identify the aberrations present at diagnosis which are driving the disease. Tracking these abnormalities in matched diagnostic

and relapse samples will identify those that persist at the point of disease progression to determine which abnormalities confer chemoresistance to the cancer.

Despite over 90% of paediatric BL and DLBCL cases being successfully cured at frontline treatment without high rates of relapse, it is hypothesised that driver aberrations are present at the point of diagnosis that allow the lymphoma to survive treatment in the small number of cases who relapse or have refractory disease. For this reason, it is important to compare multiple samples both before and after treatment where possible. In CLL it has been shown that *TP53* abnormalities at diagnosis are prognostic markers of disease progression and are actively selected for during chemotherapy in longitudinal studies of sequential biopsies from the same patient. These subclones with *TP53* aberrations were shown present at diagnosis and arose before treatment was administered (Malcikova *et al.*, 2015). Analysis of subclones showed that patients with small *TP53* abnormal clones had the same inferior outcome as those with major *TP53* aberrant clones at diagnosis (Rossi *et al.*, 2014). In CLL patients with subclonal *TP53* aberrant populations at diagnosis that subclone expanded after treatment with conventional chemotherapy. Current knowledge of clonal expansion in B-NHL is limited with no sequencing- or array-based studies conducted on paediatric B-NHL or BL.

7.2 Chapter Aims

The specific aims of the work detailed in this chapter were as follows:

- Compare the genomic landscape of relapsed/refractory paediatric B-NHL in both pre-treatment and post-treatment biopsies and track changes in profiles between both samples.
- Investigate the pattern of clonal evolution exhibited in disease progression in paediatric B-NHL.
- Determine whether genomic abnormalities are associated with patterns of clonal evolution.

7.3 Results

7.3.1 Patient Cohort

A total of 11 diagnostic and relapse pairs were analysed in this study. Nine patients from the CCLG cohort who experienced relapsed/refractory disease for whom there were both diagnostic and relapse biopsies available via the CCLG biobank. Seven CCLG cases were diagnosed BL and diagnosed were DLBCL. One DLBCL case harboured an IGH-MYC translocation while the other did not. Two additional cases were obtained from the Queen Elizabeth Central Hospital in Blantyre, Malawi. One of these cases was diagnosed BL and was analysed as part of the eBL cohort in Chapter 6, while the other was considered to be a HGBL, NOS case in line with the WHO 2016 classification (Swerdlow *et al.*, 2016).

In sBL, 4/7 relapse cases died of disease and three were alive at last follow up of 17.3, 123.1 and 45.7 months respectively (Figure 7.2). All four BL patients that died did so within 12 months of initial diagnosis at 7.1, 5.3, 11.1 and 11.1 months respectively. Of the two DLBCL cases one died of disease at 17.1 months after diagnosis while the other was still alive at 36 months.

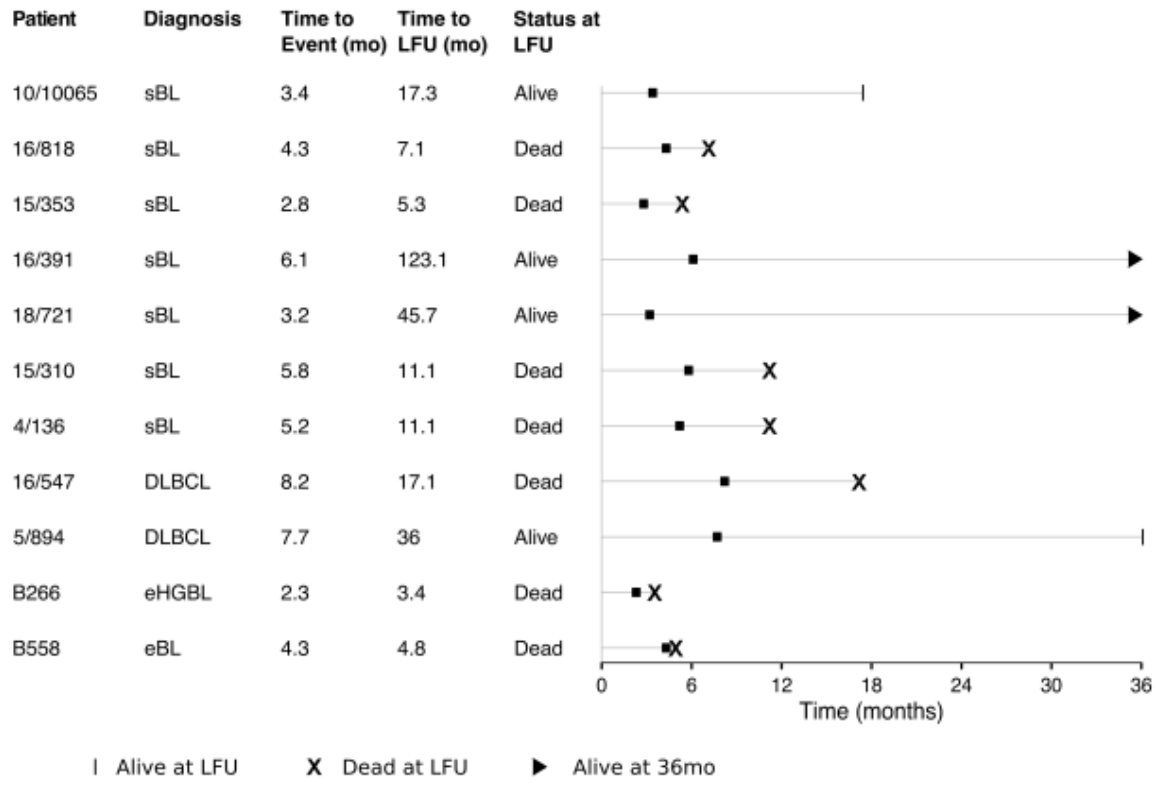


Figure 7.2 - A Forest plot displaying the clinical course for 11 relapsed/refractory B-NHL patients with copy number array data generated from paired diagnostic and sequential progressive sample.

Both DLBCL relapses occurred later than any sBL relapse and death occurred later after relapse than in sBL (relapse at 8.2 and 7.7 months, death at 17.1 months after diagnosis). Both endemic patients relapsed quickly (2.3 and 4.3 months) and death occurred in both cases rapidly thereafter (1.1 months later and 0.5 months later respectively).

7.3.2 Sporadic Burkitt lymphoma pair 1 - 10/10065

Case 10/10065 from the CCLG cohort was diagnosed as stage III BL at presentation with an IGH/MYC translocation confirmed by FISH. The 1-year old patient was male was treated on the Inter B-NHL trial with a FAB/LMB96 protocol in group B and rituximab. At the end of frontline therapy, a thymic cyst was biopsied. Relapse was confirmed within one month of remission, and the patient was alive and disease free at last follow up 17 months after diagnosis (Figure 7.2). The patient had gain of 1q, telomeric deletion of 6q, gain of the 16q telomere and deletion of 17p at diagnosis, as well as CNN-LOH affecting 1p (Figure 7.3). All copy number abnormalities identified in the diagnostic sample were present at relapse.

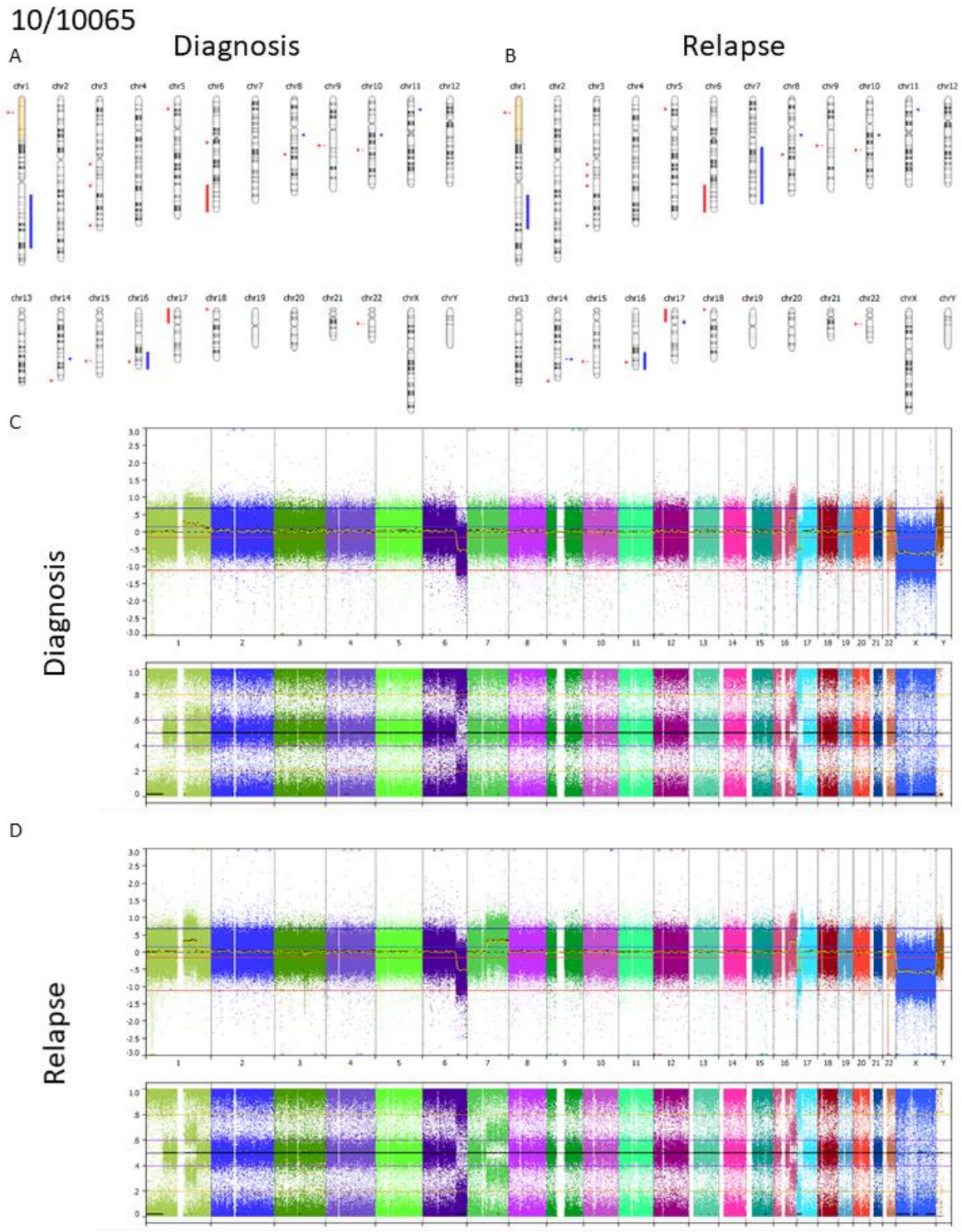


Figure 7.3 - Copy number profile of sporadic BL patient 10/10065 at diagnosis and relapse. Summary of segmented copy number aberrations at (A) diagnosis and (B) relapse. Processed log ratio data and B-allele frequency data for both (C) diagnostic and (D) relapse samples.

A gain on chromosome 1 is maintained from diagnosis to relapse (Figure 7.3A-B), accompanied by CNN-LOH on 1p and a focal deletion of both copies of *ID3* (Figure 7.4).

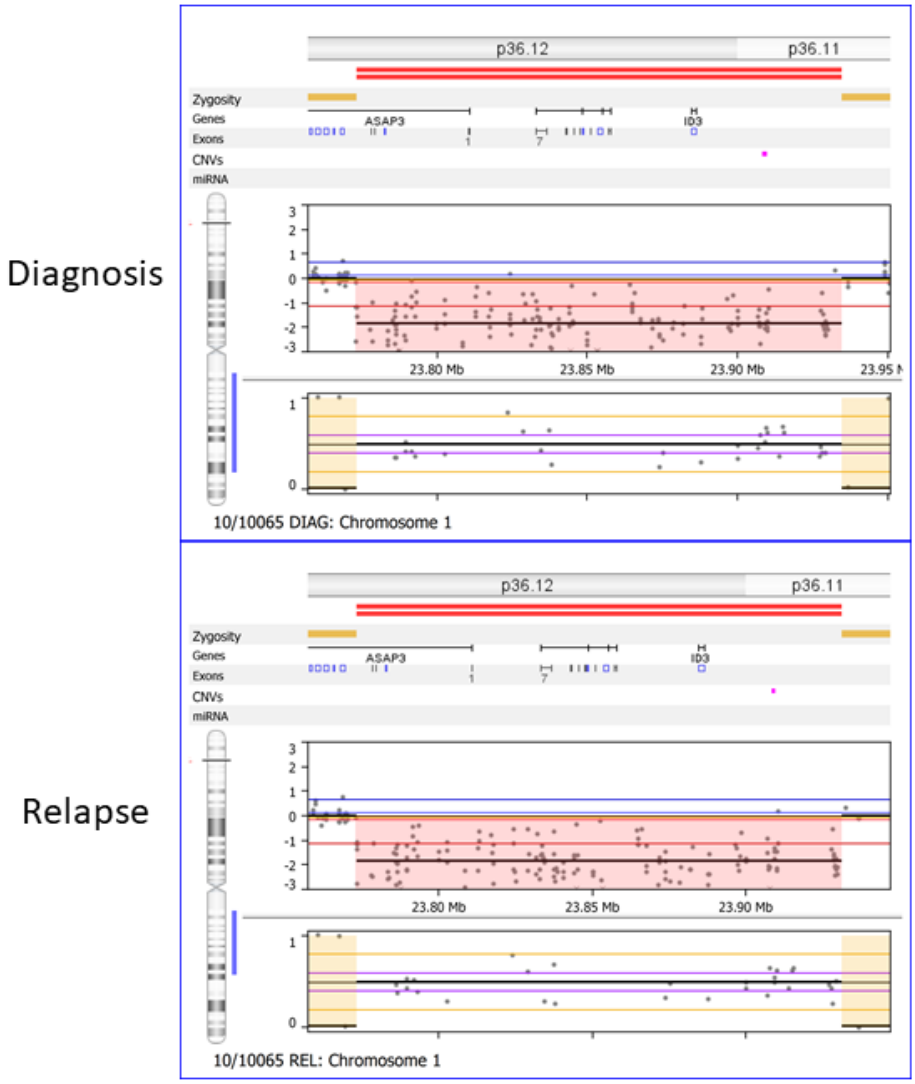


Figure 7.4 - ID3 *deletion on 1p was identified in patient samples from 10/10065 at both diagnosis and relapse.*

The breakpoint of 1q gain differed between samples, with a median lower log ratio at diagnosis suggesting a smaller population of cells harboured the aberration but spanning more of the 1q arm. 1q gain at relapse had a higher log ratio suggesting that the abnormality is present in a larger proportion of subclones within the relapse sample and the B-allele frequency plot showed a stronger separation of the central track which supports this suggestion (Figure 7.3C-D).

The relapse sample has a loss at 3q13.31 with 3 distinct steps involving the genes *TUSC7*, *LINC00903*, *LSAMP* (and *LSAMP-AS1*), *MIR4447* and *LINC00901*. The log ratio was -1.1 at the lowest point within the segment, strongly suggesting that the loss occurred in more than one subclonal population and that both copies were deleted in the majority of the tumour (Figure 7.5). The deletion of the *TUSC7* gene correlated with the 3q13.31 loss region identified by GISTIC2 in chapter 4.

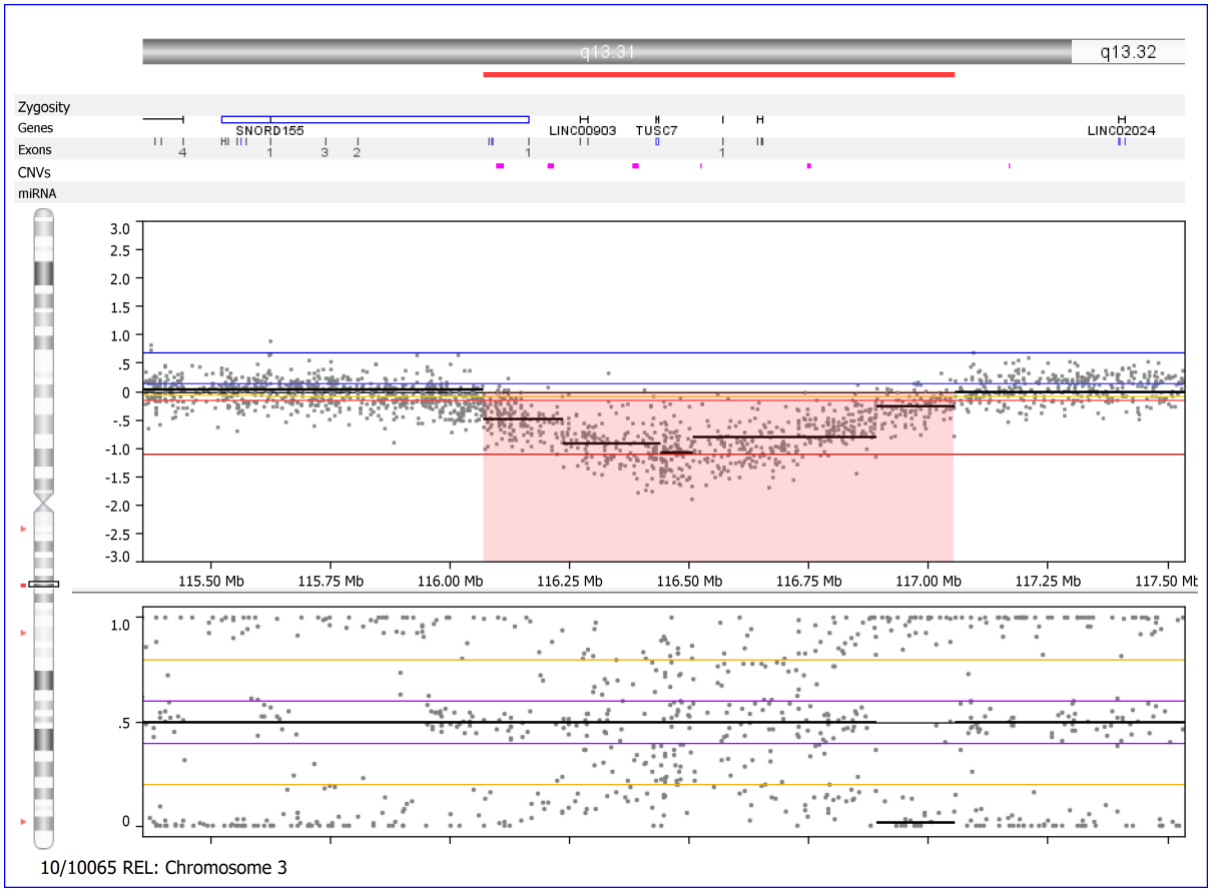


Figure 7.5 - Homozygous TUSC7 deletion at 3q13.31 was identified in the relapse sample from 10/10065 only.

Gain of 7q was observed in the relapse sample and not the diagnostic sample. 7q gain is common in B-NHL and has previously been considered a putative prognostic marker in the paediatric setting, however not in an analysis of 162 cases in chapter 4. The region gained stretched from 7q11.23 to the 7q telomere, involving the majority of the arm (Figure 7.6). There was no evidence of the abnormality being present at diagnosis, even in a small subpopulation. However, the sensitivity of copy number arrays is not sufficient to completely rule out presence of the abnormality in a very small subclonal population. Both samples had gains on 16q which had identical breakpoints and similar log ratio values between timepoints which suggests this event is stable and present in the clonal population.

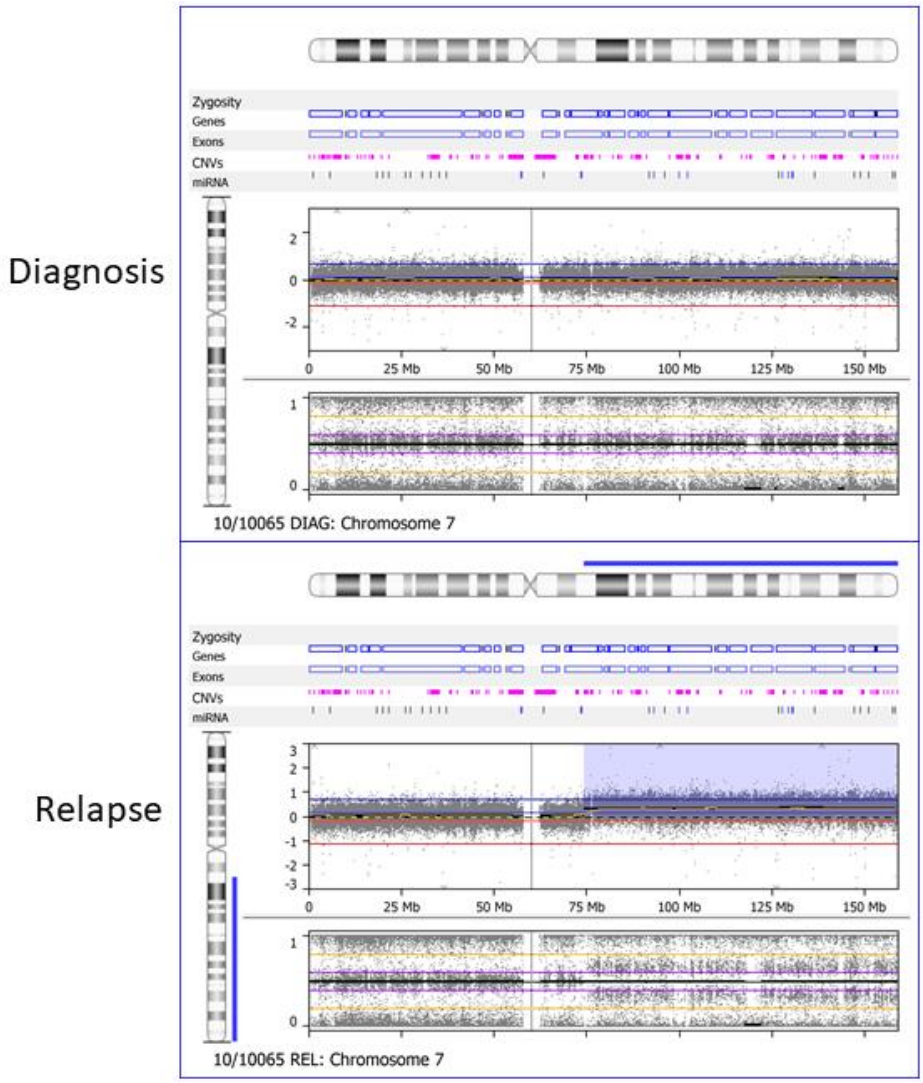


Figure 7.6 - 7q gain was acquired at relapse in patient 10/10065 but was absent at diagnosis.

The diagnostic sample of this BL patient had a heterozygous 17p deletion involving the whole arm. The arm was still deleted at the time of relapse, but a complex pattern of gain and loss was observed at the 17p centromeric region. *TP53* was still deleted despite this new pattern of copy number change arising. This event was not seen at diagnosis, even at low levels. There was no karyotyping data or constitutional/matched normal DNA sample meaning that it was not possible to screen for rearrangements with other chromosomes.

Given the early relapse experienced by this patient, it was assumed that there was not ample opportunity for extensive tumour evolution and for diverse subclonal populations to arise which would explain the overall consistency in genomic profiles between diagnosis and relapse. Specific copy number aberrations were seen that showed the genomic profile changed at relapse, suggesting that these abnormalities

were beneficial to the tumour while being treated with intensive chemotherapy. Numerous candidate genes reside on chromosome arm 7q which may confer the survival advantage to the subclone. Deletion of the *TUSC7* gene itself is a candidate tumour suppressor gene that requires further investigation.

7.3.3 Sporadic Burkitt lymphoma pair 2 – 16/818

Patient 16/818 was a 3-year-old female diagnosed with BL in the pelvis and abdomen. The malignancy was *IGH-MYC* positive by FISH but tumour stage and other clinical factors were unknown due lost clinical notes. The patient underwent an early relapse at four months after diagnosis and died of disease three months after relapse. The diagnostic sample had a quiet genome with deletions of the 1p telomere and 17p as the only large aberrations. At relapse both deletions were maintained, and multiple other copy number abnormalities were acquired. Copy number gain of 3q, the 8q telomere and chromosome 12 were observed at relapse. Two complex chromosomal aberrations involving 17q and 18q were also acquired. There was no evidence from the copy number array data that acquired aberrations were present in a subclonal population of the diagnostic sample. Oncoscan array data for the relapse sample was noisy meaning that some allelic calls were unclear during analysis, however copy number changes could be confidently called (Figure 7.7).

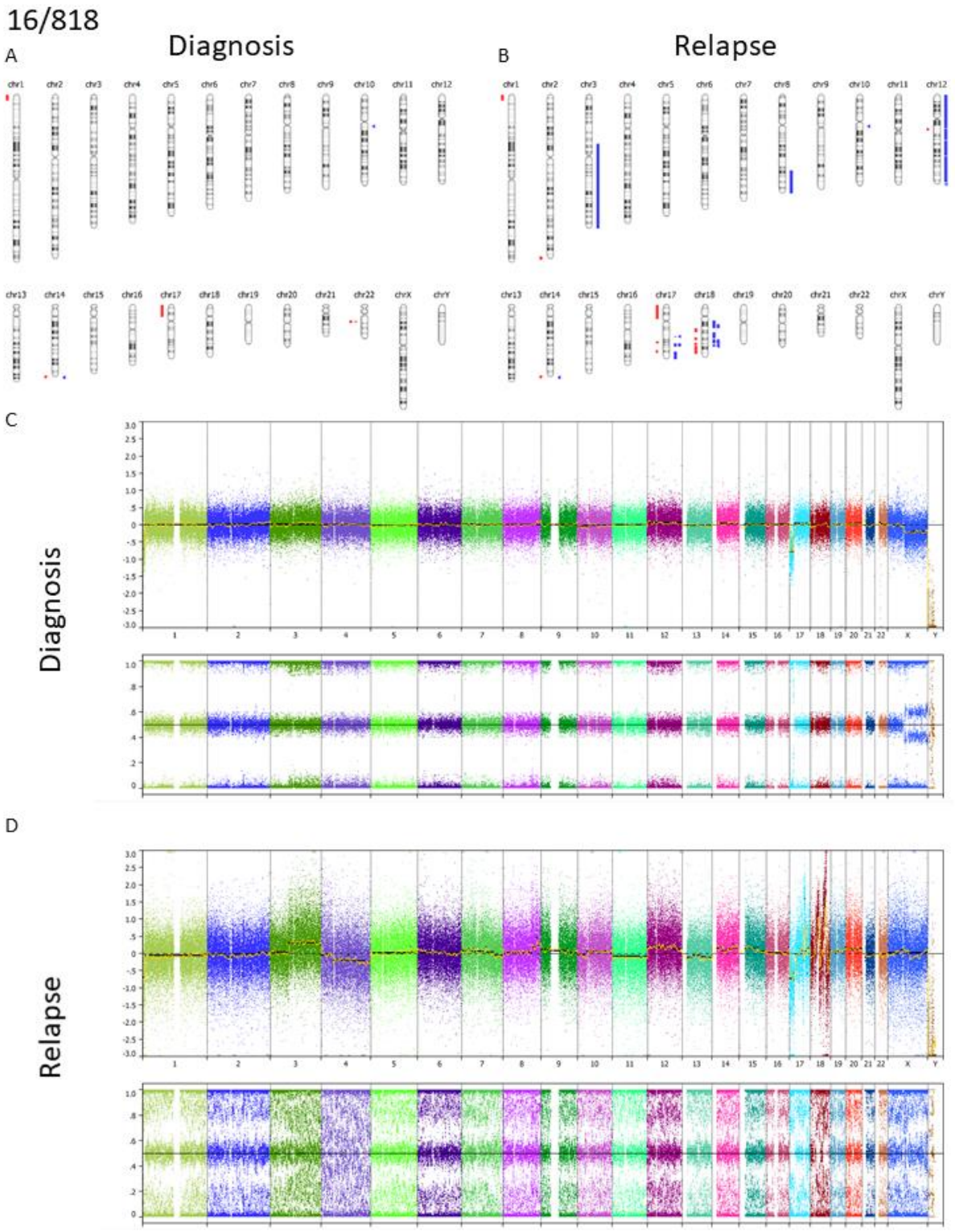


Figure 7.7 - Copy number profile of sporadic BL patient 16/818 at diagnosis and relapse. Summary of copy number aberrations at (A) diagnosis and (B) relapse. Processed log ratio data and B-allele frequency data for both (C) diagnostic and (D) relapse samples.

An 8.1Mb deletion of the 1p telomere was observed at diagnosis and maintained at relapse. The log ratio was consistently -0.69 in both samples suggesting now

expansion or contraction of the abnormality and that it was likely present in the clonal population in each case. It is likely this aberration, that involved deletion of *TP73*, occurred early on in lymphoma development. The relapse case exhibited copy number gain of the majority of chromosome 3 involving all of 3q and the centromeric region of 3p. The log ratio was 0.31 and copy number gain was not seen at all in the diagnostic sample. The region involved 3q29 which was shown to be a prognostic marker associated with relapse in chapter 4. There was no focal amplification in this region suggesting a candidate gene.

A region 33.6Mb on 8q toward the telomere stretching from 8q23.3-8q24.3 was gained in the relapse sample. Within this region of copy number gain was a smaller region with a higher copy number containing 27 genes in a 5.2Mb region at 8q24.21-8q24.22. This segment of copy number gain can be seen in a small upward shift in the log ratio in the diagnostic, but below the threshold for the segmentation algorithm to call. This supports the likelihood that a gene in this region was beneficial to lymphoma survival in this patient and that tumour cells with this aberration survived frontline chemotherapy. It is also possible that these cells had a different chemoresistance-inducing genetic alterations and that 8q gain was just a passenger aberration.

Chromosome arm 17p was lost at diagnosis and this deletion was maintained at relapse. The log ratios were similar at -0.79 and -0.72 at diagnosis and relapse respectively (Figure 7.8A-B). This suggests the deletion is clonal and occurred early in lymphomagenesis. Chromosome 17q was also aberrant in the relapse sample, but not the diagnostic. A complex pattern of gain and loss was seen on 17q including a high copy number amplification region with a log ratio of 1.61 (TuScan estimation = 8 copies) which includes *TBX2* and *TBX4*, often amplified in cancer (Figure 7.8B). *TBX2* is involved in senescence and is known to inhibit CDKN2A ($p19^{ARF}$) and abrogates the *TP53* pathway (Vance *et al.*, 2005). CNN-LOH on 17q was shown in chapter 4 to be associated with poor outcome and a higher risk of relapse but copy number changes were not common. *TBX2* and *TBX4* may be candidate genes for mutation analysis within the 17q region. Another region of amplification on 17q was seen with an estimated copy number of 6 involving several HOX family genes (*HOXB5*, *HOXB6*, *HOXB7*, *HOXB8*, *HOXB9* and *HOXB13*). Chromosome arm 18q

also harboured complex aberrations involving six regions of amplification and four regions of deletion that were specific to relapse (Figure 7.8).

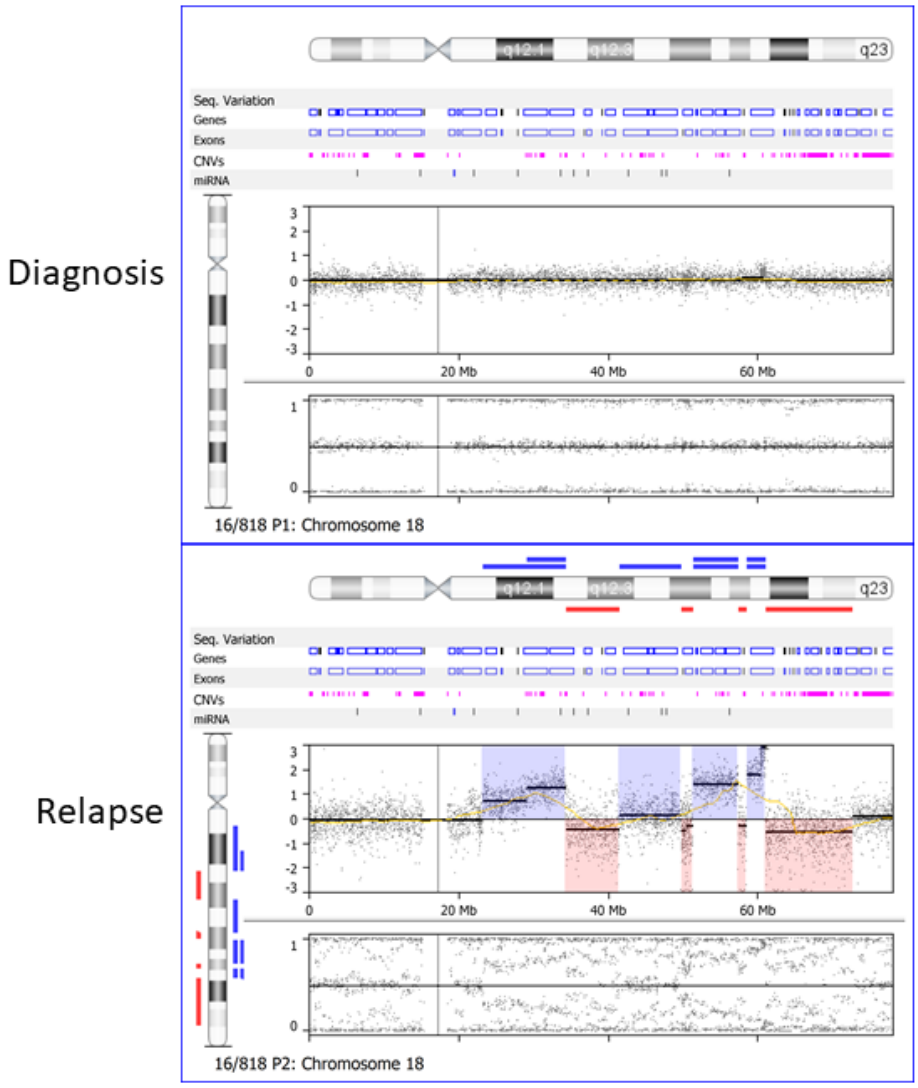


Figure 7.8 - Chromosome 18 was not aberrant at diagnosis but acquired a complex pattern of copy number aberration at relapse on the 18q arm. The aberration involved high level amplification of the BCL2 *gene with an estimated 25 copies present.*

Copy number gains occurred in a step wise pattern which culminated in the highest level of amplification with a log ratio of 2.90, estimated to be 25 copies of *BCL2* which was one of three genes in this region alongside *PHLPP1* and *KDSR*. Immediately after this segment sat a region of deletion, suggesting a complex rearrangement similar to those associated with *TP53* aberration described in chapter 5.

The pattern of clonal evolution in this case of early relapse resembled that of 10/10065 in that the diagnostic and relapse samples were closely related genomically. Abnormalities on chromosome 12, 17q and 18q arose at relapse

presenting putative candidate genes that are known to play a role in lymphoma that may be responsible for driving disease progression. Some evidence of subclonal populations was detected on chromosome 8 and chromosome 12, suggesting that the resistant clone had expanded and given rise to further subclones with biallelic inactivation of *TP53*, 17q and 18q aberrations in common.

7.3.4 Sporadic Burkitt lymphoma pair 3 – 15/353

Patient 15/353 was diagnosed with stage III *IGH/MYC*-positive BL at the age of 11. The patient was male and masses in the abdomen and mediastinum and there was no involvement in the bone marrow or central nervous system. He was treated on FAB/LMB96 protocols in group B the patient relapsed one month after final frontline chemotherapy and died of disease three months after relapse. At diagnosis the patient had multiple small aberrations as well as 6p CNN-LOH that was maintained at relapse. The relapse sample acquired a complex pattern of copy number gain and CNN-LOH on chromosome 1 and a complex pattern of gain and loss on 3q (Figure 7.9).

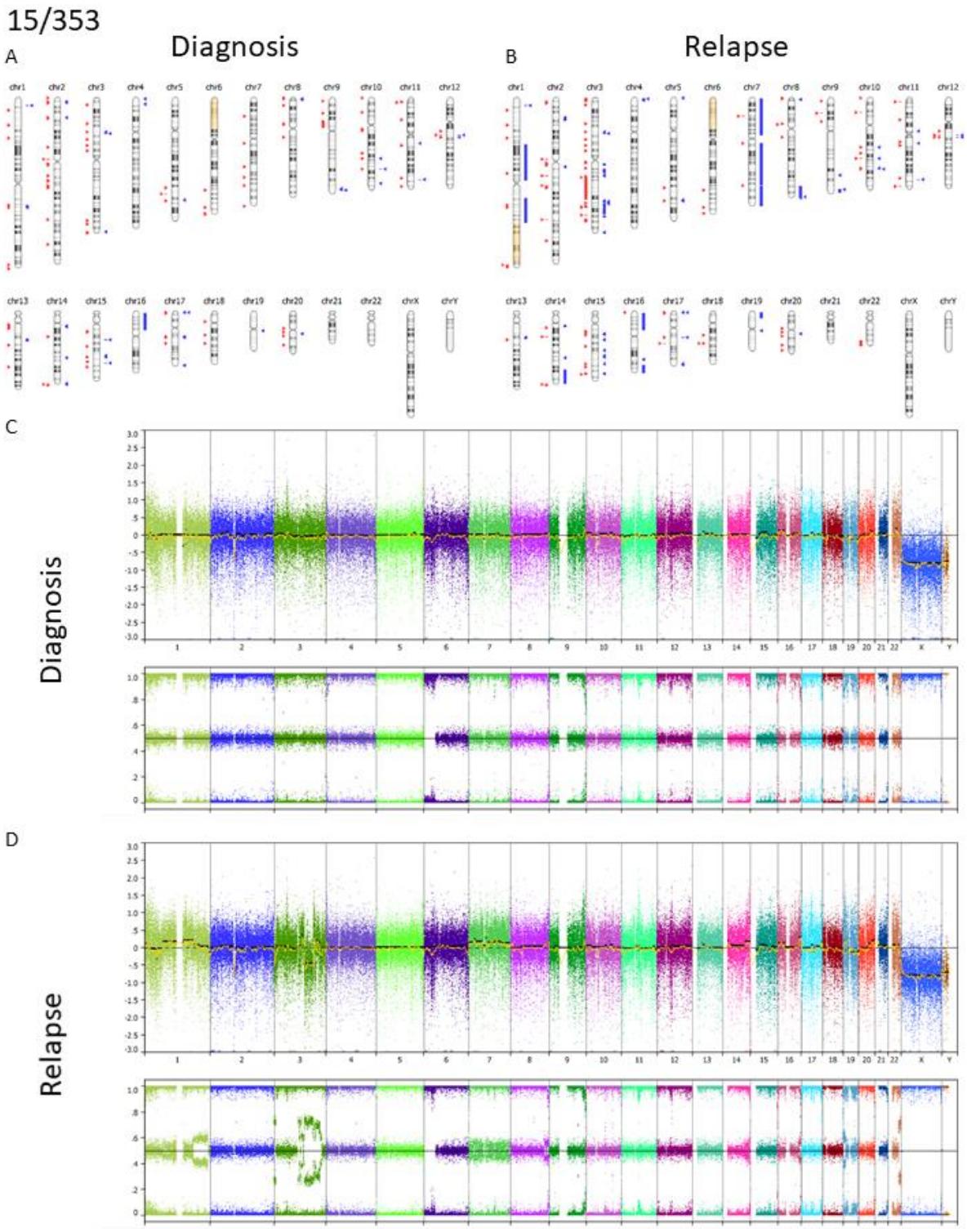


Figure 7.9 - Copy number profile of sporadic BL patient 15/353 at diagnosis and relapse. Summary of segmented copy number aberrations at (A) diagnosis and (B) relapse. Processed log ratio data and B-allele frequency data for both (C) diagnostic and (D) relapse samples.

Gain of the chromosome 1 centromere was seen at relapse but not diagnosis (Figure 7.9), involving the majority of 1q and 52.5Mb of 1p. Gain on 1q was paired with telomeric loss and there were two distinct copy number segments in 1q. The region adjacent to the centromere appeared to have gained two additional copies while the

rest of the arm had one additional copy with the exception of the telomere which was deleted. Aberration within the *ALK* gene in both samples was observed on chromosome 2. Both cases had deletion of exon 1 of the gene followed by gain of an additional copy of exons 2-11, followed again by deletion of exon 12-15. The remaining exons were unaffected, and breakpoints were consistent between samples. The *ALK* gene is a recurrent translocation partner in rearrangements in anaplastic large cell lymphoma (ALCL) and encodes a tyrosine kinase receptor (Ferreri *et al.*, 2012). The diagnostic sample harboured a focal 3q29 gain involving only the *TFRC* gene (transferrin receptor 1) or CD71 which is highly expressed in BL and expression is a believed to be a poor prognostic marker in follicular lymphoma (FL) treated with rituximab (de Fontbrune, 2019). The remainder of chromosome 3q was unaffected by copy number aberrations in the diagnostic sample but was highly complex in the relapse sample. A complex pattern of 10 gain and 10 loss segments spanning 87.8Mb stretched from 3q11.1 to 3q26.33 (Figure 7.10). Within this region was a small focal homozygous deletion involving the *TUSC7* gene, as was seen in 10/10065 at relapse. The focal gain of *TFRC* was maintained at relapse.

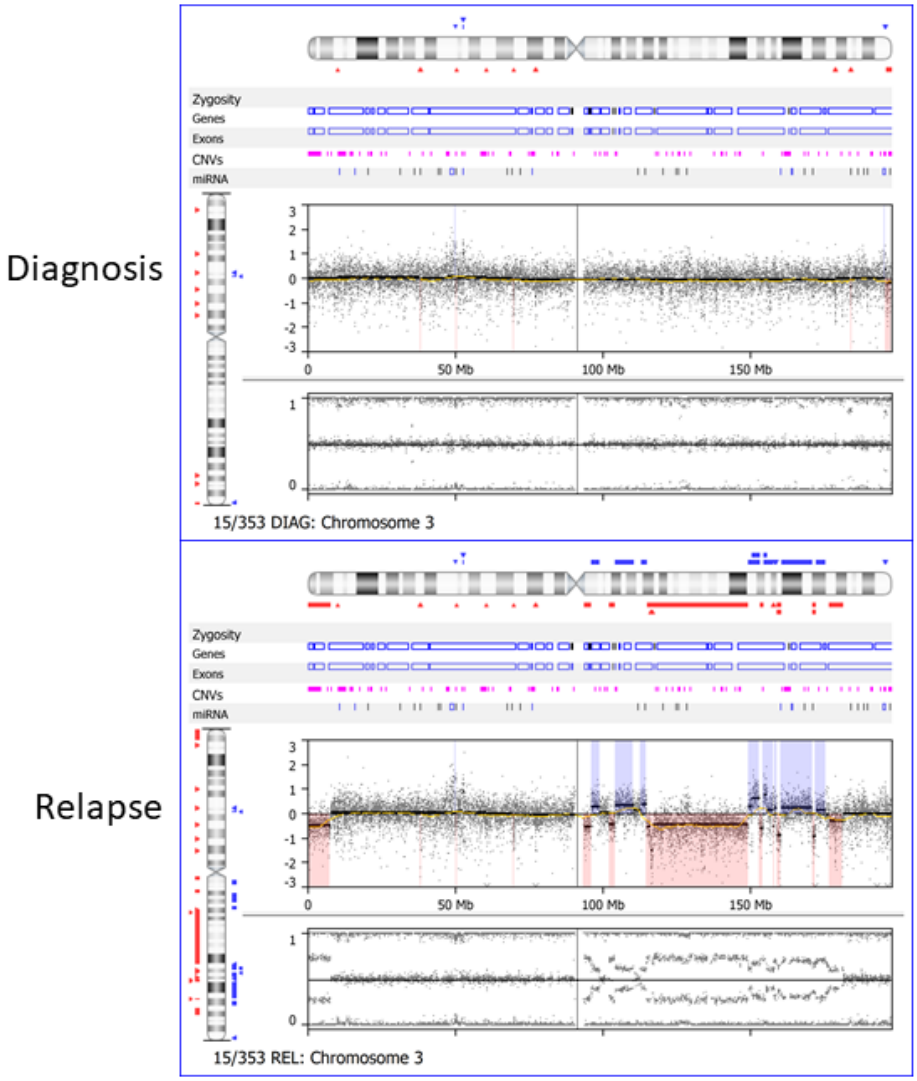


Figure 7.10 - Complex aberration on 3q involving multiple regions of copy number gain and loss was observed in the relapse sample of 15/353 that was not present in the diagnostic biopsy.

A low-level gain of chromosome 7 was seen at relapse which could not have been present in the major clone as the log ratio was low at 0.17 and the was no evidence of this aberration in the diagnostic sample. The relapse sample had an additional copy number gain at 8q24.22 to 8q24.3. The breakpoint furthest from the telomere was adjacent to the *MYC* gene, which was found within a focal deletion involving only the gene.

Two deletions on 9p at 9p24.1 and 9p21.3 were present at diagnosis involving the *JAK2* and *CDKN2A* genes. Both deletions were present at relapse with similar log ratios in each case (*JAK2*: -0.86 vs -0.62, *CDKN2A*: -1.13 vs -0.92) suggesting that the aberrations occurred early during the development of the malignancy. *CDKN2A* deletions in both samples were entirely focal involved no other genes (Figure 7.11).

In addition, exon 1 of *PTEN* was deleted in both samples on chromosome 10 and focal amplification of the *CBL* gene on 11q was maintained at relapse.

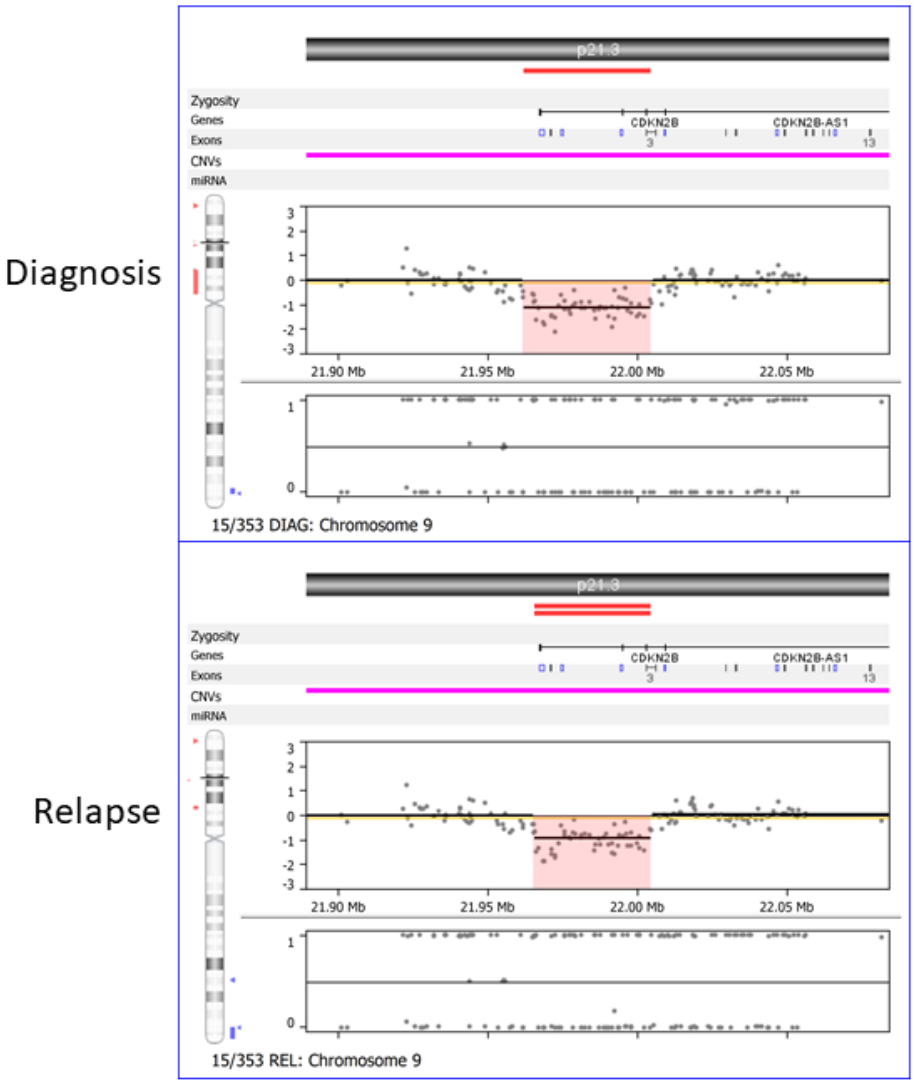


Figure 7.11 - Partial CDKN2A deletion identified at diagnosis and maintained at relapse in patient 15/353. Region shown spanned chr9:21889358-22084295.

A region of copy number gain on 14q telomere was seen in the relapse sample that was present in the diagnostic sample but involving copy number gain of more of the chromosome. The breakpoint region of the gain on 14q sat within the *MYC* gene which may suggest that *MYC* rearrangement in this case involved an imbalanced translocation that is more complex than those typically found in BL. No karyotyping data was available for either sample and no constitutional DNA was present for structural analysis by whole genome sequencing. 16p gain was seen in both samples with the same breakpoints, but a 11.3Mb gain on 16q at the telomere was seen only at relapse. Chromosome 19p gain was seen only in the relapse sample, affecting 19p13.3 to 19p13.2 involving a region identified by GISTIC2.0 analysis in Chapter 4.

Clonal evolution in this relapse case resembled that of both 10/10065 and 16/818 as the two samples were closely related in their genomic profiles. The additional abnormalities arising at the point of relapse that could not be identified as subclonal aberrations at diagnosis suggest that genomic instability is increased as disease progresses.

7.3.5 Sporadic Burkitt lymphoma pair 4 – 16/391

Patient 16/391 was a 13-year-old male diagnosed with MYC positive BL. The tumour had no bone marrow or central nervous system involvement and was diagnosed an abdominal stage III BL. After successful treatment on FAB/LMB96 protocols in group B the patient relapsed 3 months on from remission. The patient remains alive and disease-free 10 years on from diagnosis. The diagnostic tumour had copy number gains of 1q, chromosome 7, chromosome 9, chromosome 19 and 21q, and a complex aberration on chromosome 11q, typical of those seen in BLL-11q cases. The relapse sample acquired an additional copy of chromosome 12 as well as complex chromosomal aberrations of 14q and 18q. 4q deletion was also acquired, and 9q gain was not seen at relapse. Both diagnostic and relapse samples had CNN-LOH across the whole of chromosome 17.

A stepwise pattern of gain on chromosome 1q was seen in the diagnostic sample and was maintained at relapse with the same breakpoint. The highest point of gain was seen at 1q25.2 involving nine genes, none of which are known to be involved in cancer to date (Figure 7.12).

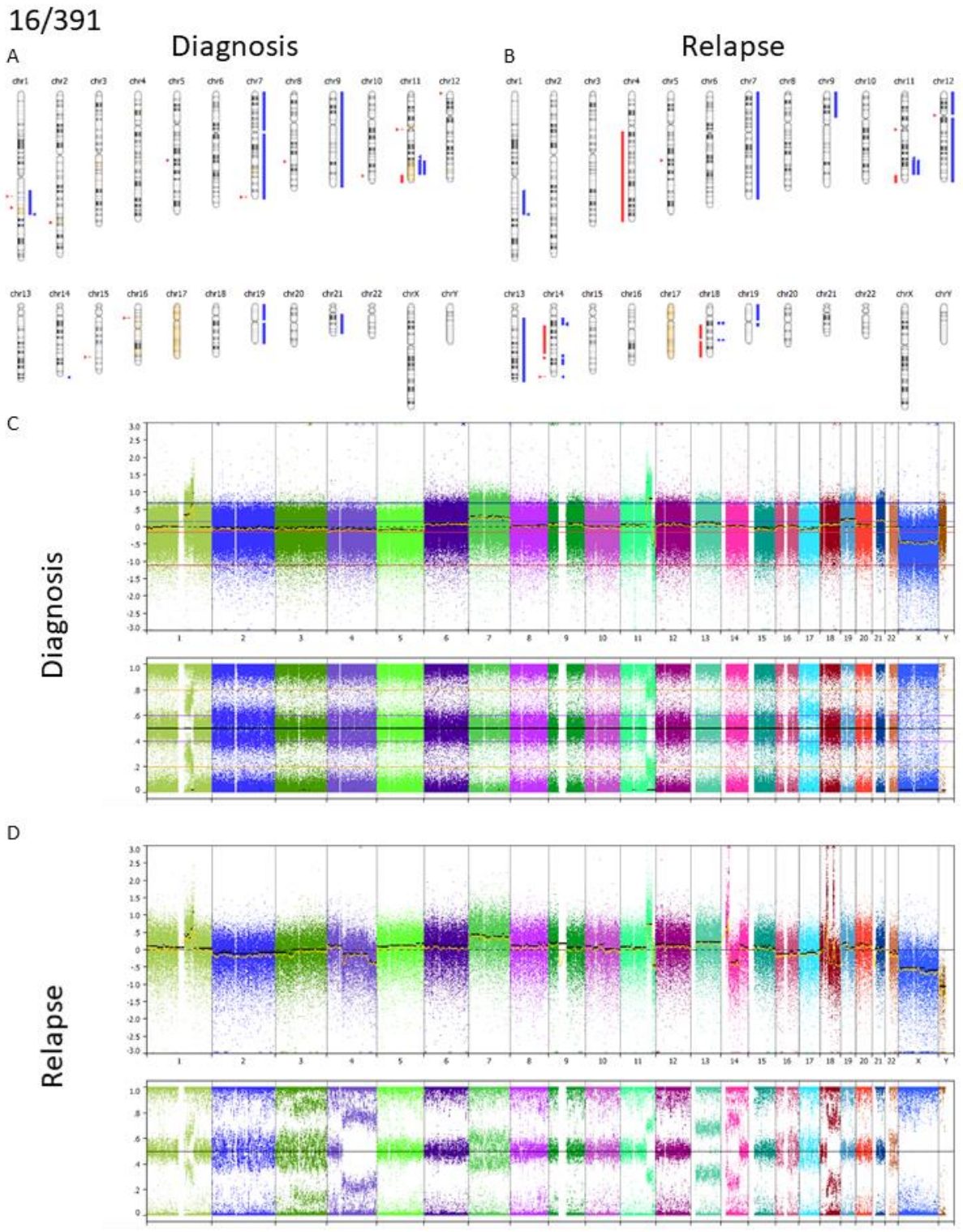


Figure 7.12 - Copy number profile of sporadic BL patient 16/391 at diagnosis and relapse. Summary of segmented copy number aberrations at (A) diagnosis and (B) relapse. Processed log ratio data and B-allele frequency data for both (C) diagnostic and (D) relapse samples.

4q deletion was not present at diagnosis but was observed at relapse spanning 4q12 to 4q35.2. Chromosome 7 was gained at both diagnosis and relapse. The log ratio remained effectively unchanged throughout the two timepoints (Diagnosis 0.40,

Relapse 0.39). A subclonal population at diagnosis had an additional copy of chromosome 9 with a log ratio of 0.14. At relapse 9p gain was still observed with a slightly higher log ratio, but 9q was not gained or altered at all. It is possible that 9p and 9q gains were present within different subpopulations at diagnosis and that the 9q gained clone was lost during treatment (Figure 7.13).

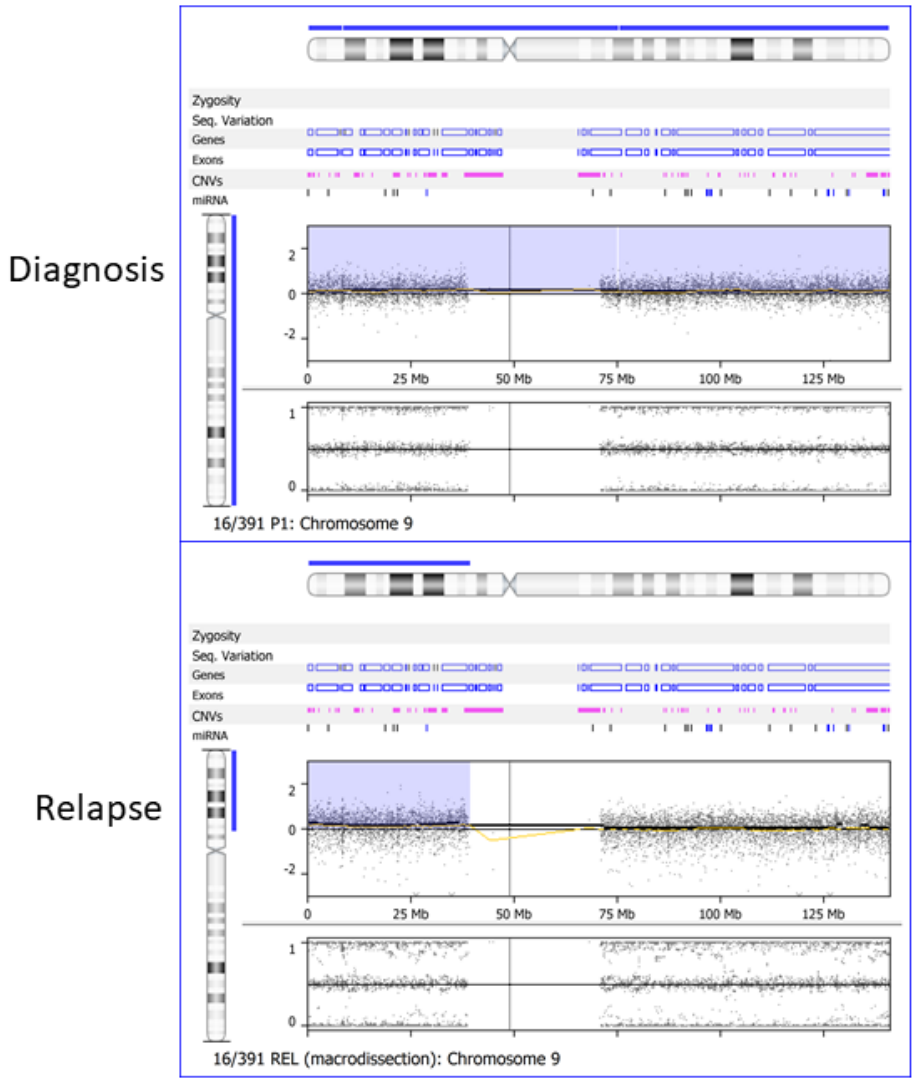


Figure 7.13 - Chromosome 9 gain at diagnosis and relapse in patient 16/391. Aberrations in both samples were barely detectable and likely affected a small subclone in each biopsy.

The copy number gains in both samples were almost undetectable but comparison of the median log ratio of probes across chromosome 9 in each sample against other chromosomes shows a clear upward shift in intensity (Figure 7.12C-D).

Chromosome 11q was aberrant at diagnosis in a pattern typical of BLL-11q cases described recently (Salaverria *et al.*, 2014). The gain spanned 22.2Mb at 11q22.1-

11q24.1 but a peak of copy number gain was observed at 11q22.1-11q22.3 involving 18 genes including multiple matrix metalloproteases (MMP) genes as well as *BIRC2* and *BIRC3* which both play a role in TNF signalling and resistance to apoptosis (Figure 7.14). The breakpoint between gain and loss sits within the *JHY* gene. The log ratio of gain was higher in the diagnostic sample which was likely a result of the relapse sample being small and having to be macrodissected.

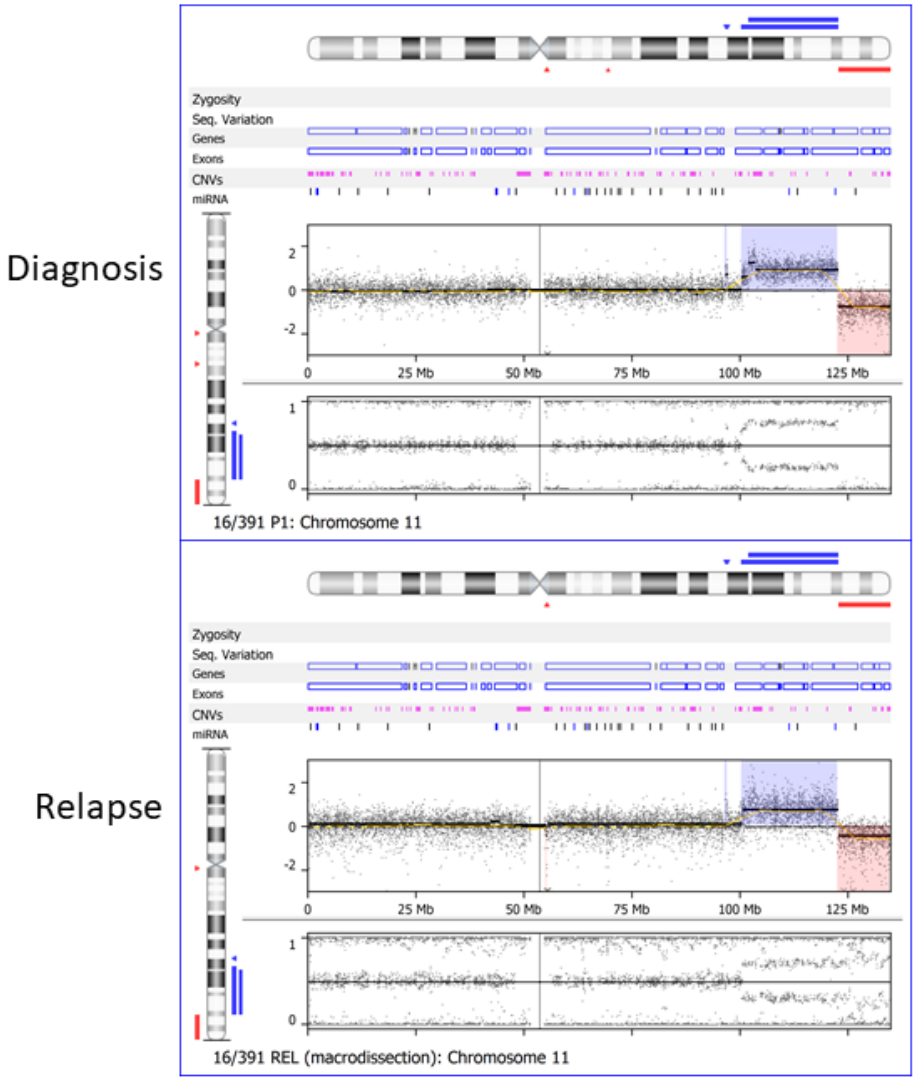


Figure 7.14 - Chromosome 11q abnormality at diagnosis and relapse in patient 16/391.

At relapse the tumour acquired a low-level gain of chromosome 12. This was considered to be a subclonal event as the log ratio of 0.13 was considerably lower than that of the gain on chromosome 1q (log ratio 0.38). Gain on 13q was likely subclonal also with a log ratio of 0.21, but in a larger subclonal population of cells. There was a complex pattern of gain and loss on chromosome 14 in the relapse

sample which involved a high copy amplification of a 1.3Mb region within 14q12 involving G2E3 (G2/M phase specific E3 ubiquitin ligase).

Chromosome 17 was aberrant in both diagnostic and relapse samples from the patient. The diagnostic tumour sample had CNN-LOH across both 17p and 17q. The BAF showed that the CNN-LOH was not present in all of the tumour sample and may have been a subclonal event at both diagnosis and relapse. However due to the macrodissected nature of the relapse sample it cannot be ruled out that the 17p CNN-LOH-bearing clone expanded into the niche left by chemosensitive clones yet was diluted in array analysis by contaminating normal tissue.

Complex aberrations on chromosome 18 involved almost the entirety of the 18q arm and was comprised of two regions of high copy gain separated by two regions of loss. Unlike the 18q complexity seen at relapse in 16/818, this aberration did not involve gain of *BCL2*. The *TCF4* and *POLI* genes were found in the regions of high copy number gain. A subclonal gain of chromosome 19 was seen at diagnosis which is present in part at relapse. In the relapse sample the abnormality was also subclonal but did not involve the majority of 19q. Similar to the aberration on chromosome 9 it is likely that there were multiple chromosome 19 events that occurred in separate subclones, some of which were not maintained at relapse. Chromosome 21 gain was seen at diagnosis but not at relapse. The gain was likely subclonal with a log ratio of 0.21. The B allele frequency track separated by a small amount on chromosome 22 in the relapse sample which suggests allelic imbalance. No change in log ratio was observed however, suggesting that instead of a subclonal copy number gain or loss it was more likely caused by a subclonal CNN-LOH event.

16/391 followed the trend of other BL relapse pairs in that copy number events at diagnosis were mostly maintained at diagnosis while new aberrations arose at relapse from subclonal populations.

7.3.6 Sporadic Burkitt lymphoma pair 5 – 18/721

Patient 18/721 was a 6-year-old male patient diagnosed with BL with an IGH-MYC rearrangement. The patient presented with an abdominal mass and was negative for bone marrow and CNS involvement. The patient was treated with FAB/LMB-based treatment protocols but treatment group was unknown. There was residual active

disease at the end of frontline therapy signalling primary refractory disease. Despite this the patient was still alive three years after the end of treatment. Diagnostic biopsy material was present in the form of an FFPE block only, while relapse biopsy material was split into one fresh frozen mass and one FFPE block with the same pathology number. All three samples were investigated by copy number array. Additional constitutional DNA was available and analysed by array. The genome of the diagnostic sample was relatively quiet with CNN-LOH on 2p and 17q only (Figure 7.15, Figure 7.16). The relapse samples harboured more aberrations and were not identical to each other. 2p and 17q CNN-LOH were not seen at relapse.

Chromosome 1 was aberrant at diagnosis and relapse, with a 1pter deletion that was maintained as well as a focal deletion of *ID3*. A small low-level copy number gain at 1q22-1q23.3 was seen at diagnosis falling within the region called by GISTIC2.0 analysis in chapter 4, but at diagnosis the aberrations on 1q were much more complex with a stepwise gain of 1q with 13 distinct segments of copy number change (Figure 7.15A-C, Figure 7.16A-C, Figure 7.17A-C).

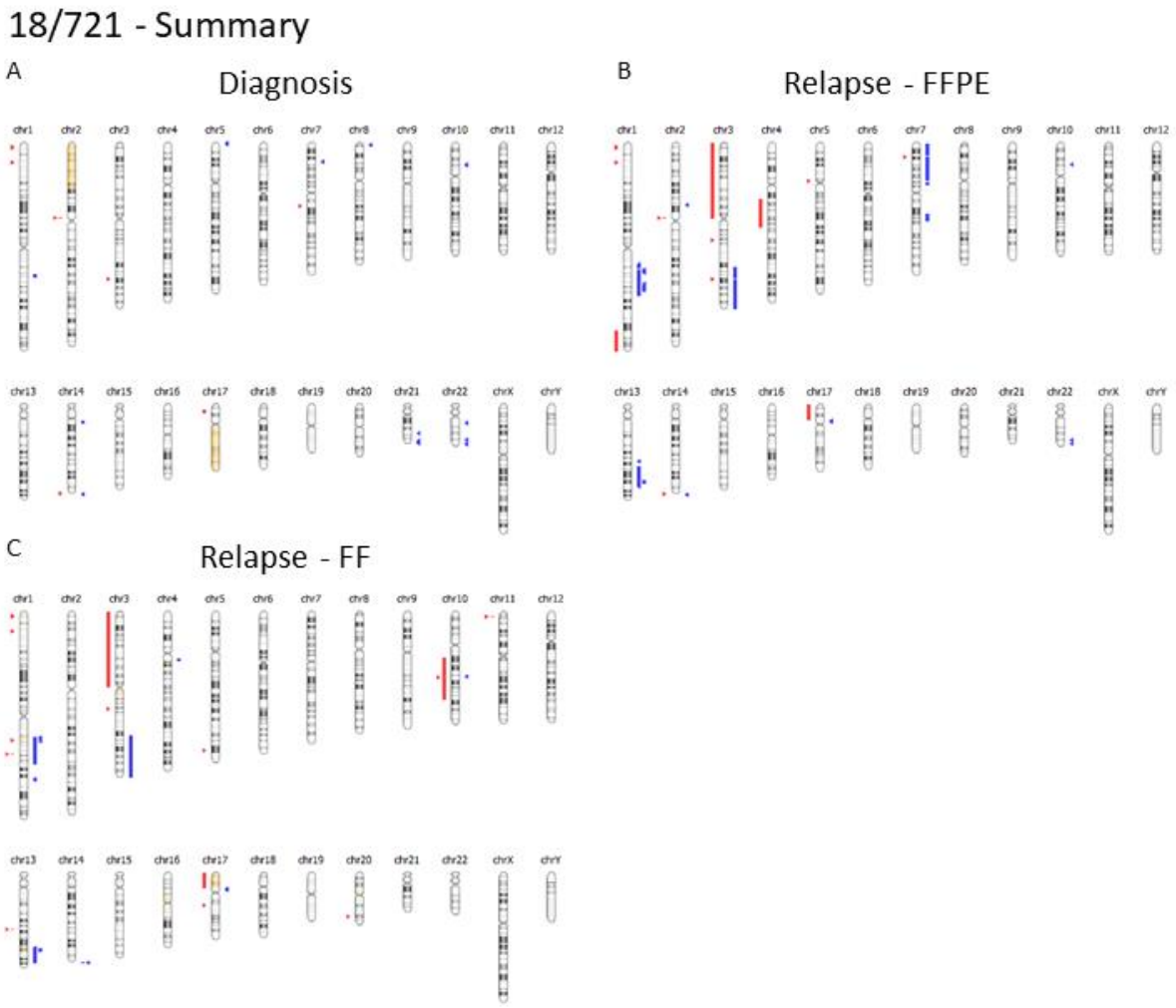


Figure 7.15 - Copy number profile of sporadic BL patient 18/721 at diagnosis and relapse. Summary of segmented copy number aberrations at (A) diagnosis and relapse in (B) FFPE and (C) FF samples from the same biopsy.

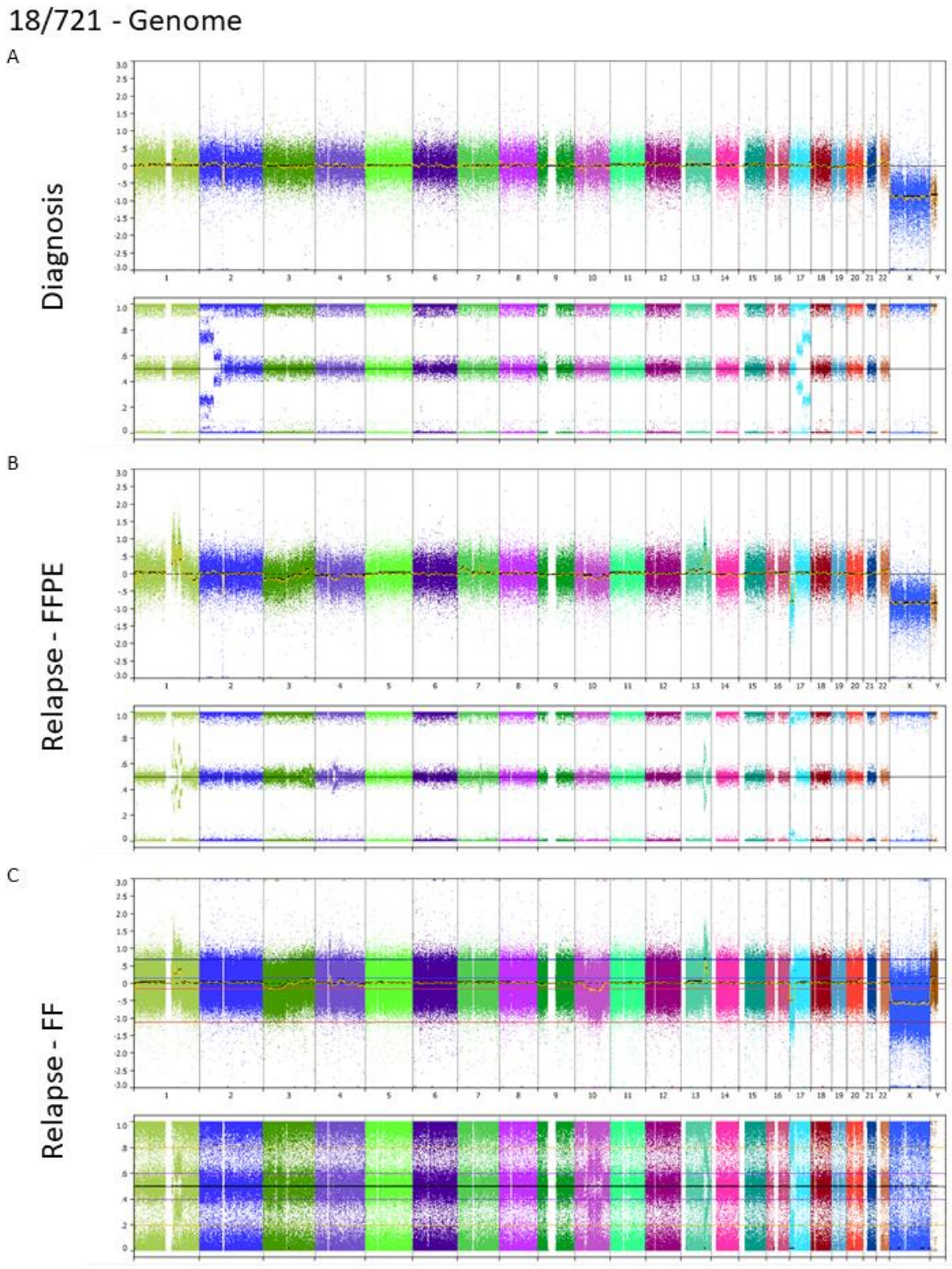


Figure 7.16 - Processed log ratio data and B-allele frequency data for (A) the diagnostic sample and relapse (B) FFPE sample and (C) FF sample from the same biopsy from 18/721.

The telomere of 1q was deleted also, but likely only in a subclone as the log ratio was only -0.16 in the FFPE relapse sample. Strikingly, the loss was not observed in the FF relapse sample. This difference between arrays represents spatial heterogeneity

within the tumour sample where different subclonal populations likely made up varying proportions of the tumour cells in each sample.

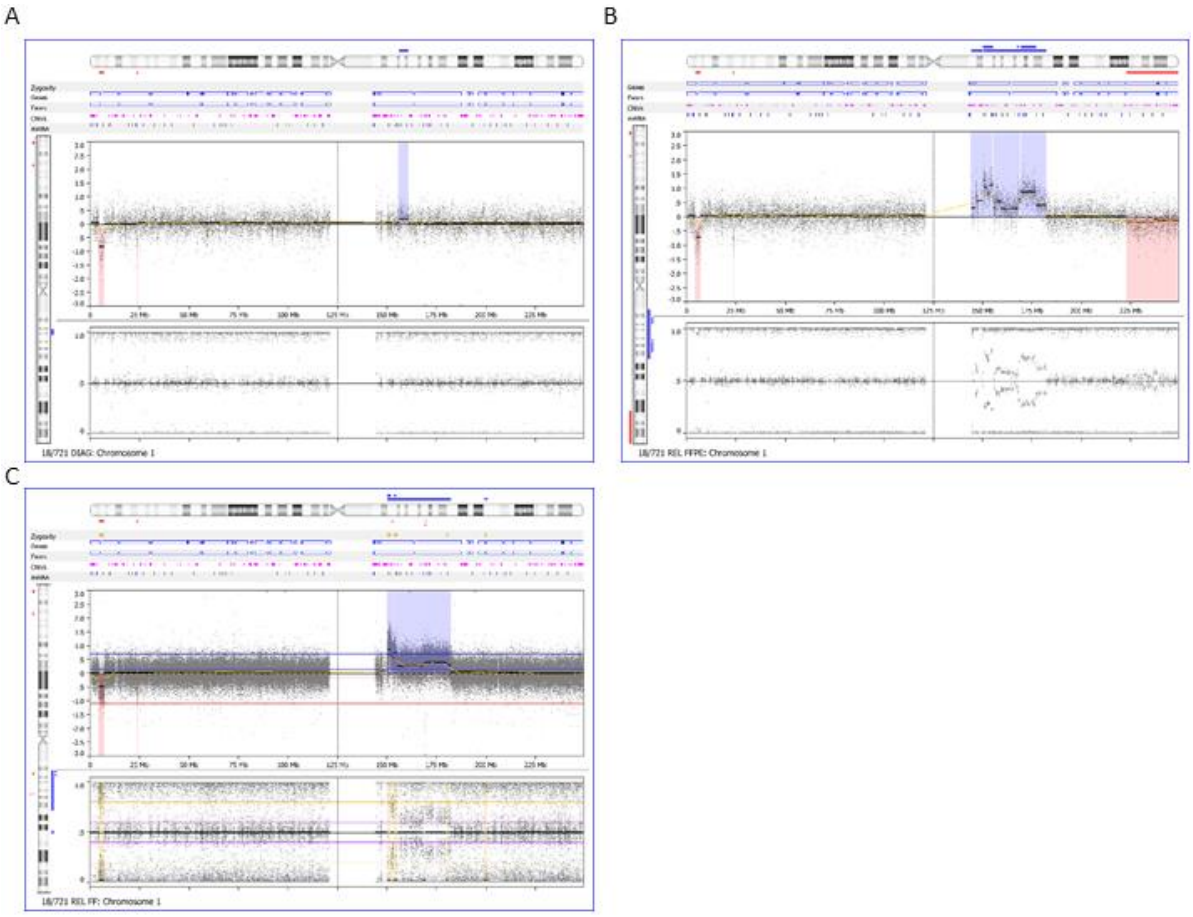


Figure 7.17 - Log ratio and B-allele frequency data for chromosome 1 depicting a complex pattern of copy number gain and amplification at 1q acquired at relapse. The abnormality was not present at (A) diagnosis but present in both (B) FFPE and (C) FF relapse samples taken from 18/721.

At diagnosis chromosome 2p was shown to have a subclonal CNN-LOH event highlighted by the four tracks observed in the BAF track. There were two distinct segments of adjoined CNN-LOH that had different BAF profiles suggesting that while both were subclonal, they were present at different proportions at diagnosis. The CNN-LOH aberrations were not seen in either relapse sample, implying that 2p CNN-LOH is not advantageous to the tumour cell in a therapeutic setting. On chromosome 3 the diagnostic sample was quiet with no copy number changes, but both relapse samples acquired low level loss of 3p and gain of 3q24 to 3q29. These segments indicated a low level event in a subclonal population as the log ratio did not deviate strongly from the baseline (gain log ratio 0.12 FF, 0.15 FFPE, loss log ratio 0.12 FF, 0.15 FFPE). In contrast, a focal deletion of *TUSC7* at 3q13.31 was most likely clonal with a log ratio of -0.73. Intriguingly the FFPE relapse sample showed a subclonal

deletion in the middle of 4q while the FF sample had a high copy gain near the centromere, with neither sample harbouring both abnormalities. This further implies a more complex tumour heterogeneity at relapse.

Chromosome 7 was quiet at diagnosis but gain of 7p was seen at relapse, but only in the FFPE sample. The 7p gain was subclonal with a log ratio of 0.13, and a small gain on 7q21.11-7q21.3 which appeared to be present at a higher incidence within the tumour at 0.28 log ratio. Chromosome 10 held another aberration that was seen in one relapse sample but not the other at 10q21.1-10q25.1 which was deleted in the FF relapse sample only. The region involved the *PTEN* gene.

Chromosome 13 was aberrant in both relapse samples while the diagnostic sample harboured no aberrations. Both relapse samples had stepwise gain with a peak segment at 13q31.3 that was common between both arrays (Figure 7.18A-B).

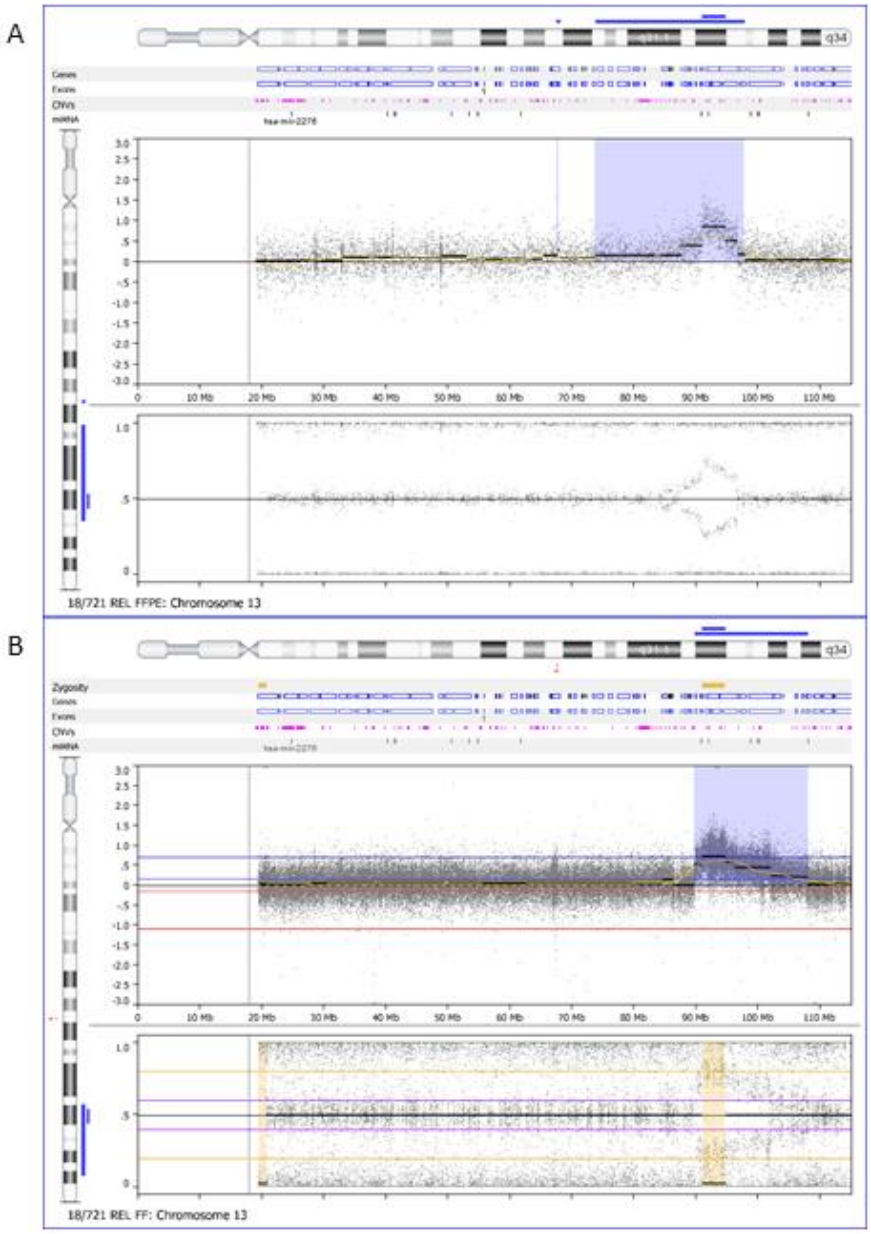


Figure 7.18 - Log ratio and B-allele frequency data for chromosome 13q depicting a complex copy number gain acquired at relapse in the (A) FFPE sample and (B) FF sample from 18/721. Both samples have amplification of 13q but the patterns are distinct.

The pattern of gain was different in both samples: the FF sample started from a high copy number gain at 13q31.3 that tailed off with segments of reducing log ratio toward the telomere, while in the FFPE sample there were multiple stepwise segments of gain on either side of the peak. It is likely that the two relapse samples represent multiple occurrences of 13q gain, with a common region of amplification involving the *MIR17HG* cluster gene.

Case 18/721 was distinct from other BL relapse pairs analysed in that the diagnostic sample bore little resemblance to the relapse sample. Only a deletion on 1p was maintained from diagnosis to relapse which suggests that this was an aberration that occurred early in the development of lymphoma. The dominant clone at diagnosis was eliminated by frontline therapy, but multiple resistant clones arose with little in common, alluded to by the variety of copy number profiles seen between the FF and FFPE relapse samples. These clones may have only had 13q gain and *TP53* biallelic inactivation in common. 18/721 experienced disease progression quicker than all but one BL relapse pairs, which is remarkable as it is generally assumed that the quicker relapses seen in paediatric BL patients lead to a linear model of clonal evolution. In this model the relapse/refractory tumour strongly resembles the tumour at diagnosis, which accurately describes BL pairs 1-4. This case however shows evidence of an later divergence of BL clones which is the alternative model recognised in B-NHL (Juskevicius *et al.*, 2017), despite there being very little time for the heterogeneity that was observed to arise.

7.3.7 Sporadic Burkitt lymphoma pair 6 – 15/310

Patient 15/310 was a 10-year-old male diagnosed with BL in multiple sites. There was no evidence of bone marrow or CNS involvement and IG-MYC translocation was confirmed by FISH. The patient was treated with FAB/LMB-based protocols in group B and was in remission at the end of CYM1. Relapse was confirmed three months after the end of treatment, or 5.8 months after diagnosis. Salvage therapy at relapse was ultimately unsuccessful and the patient died 5 months after relapse was confirmed.

A deletion of 1p35.3-1p35.1 was present at diagnosis but not observed in the relapse sample. Allelic frequency data from the array indicated the region was hemizygous and that the deletion was clonal (Figure 7.19A-D). The relapse sample had an additional copy of chromosome 7 in a small subclonal population that was not observed at diagnosis. The log ratio was very low at 0.09 but is clearly discernible when compared to the rest of the genome at relapse (Figure 7.19C-D). At diagnosis the patient had 13q CNN-LOH affecting the entirety of the arm affecting the clonal population of the tumour. At relapse 13q was entirely normal.

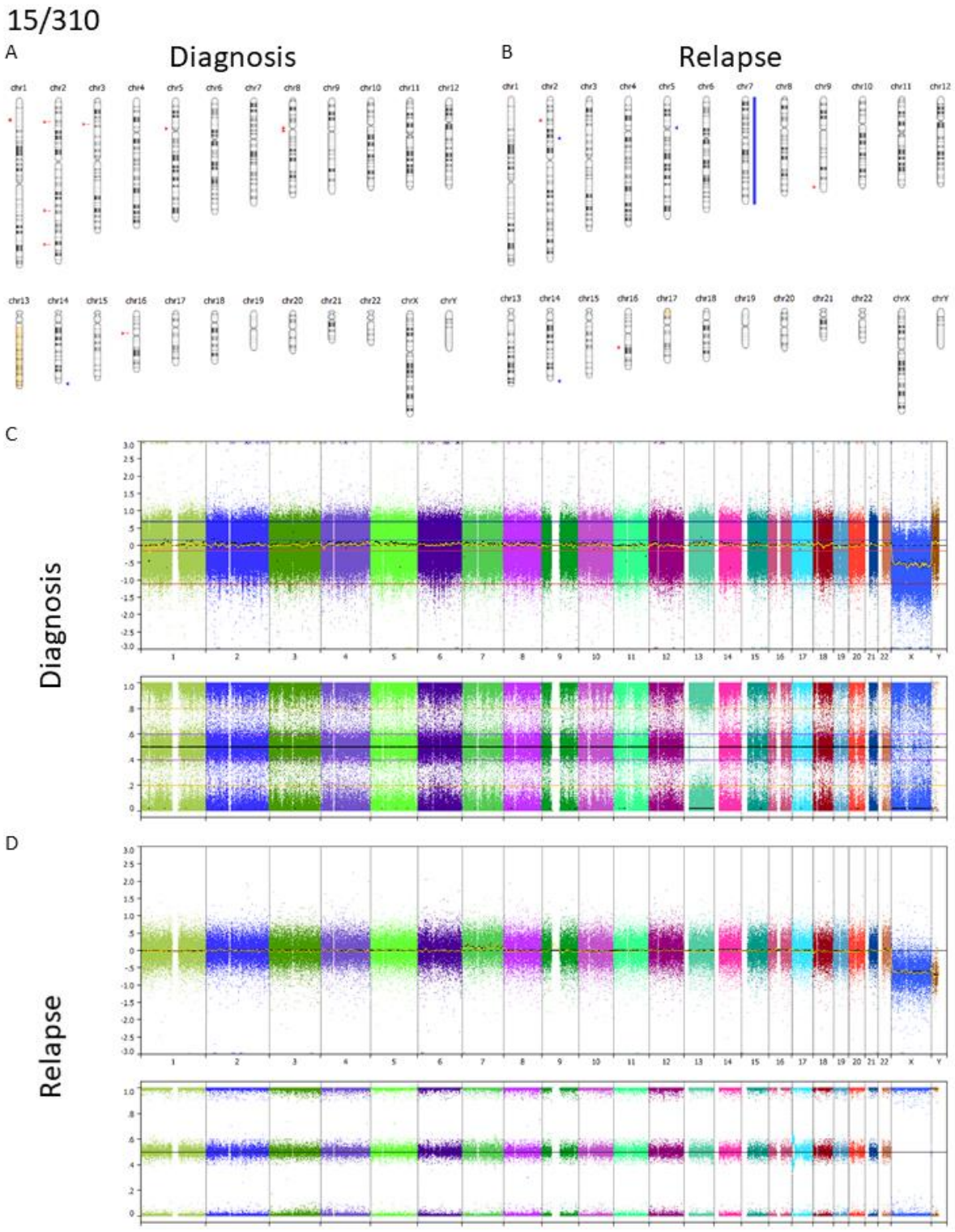


Figure 7.19 - Copy number profile of sporadic BL patient 15/310 at diagnosis and relapse. Summary of segmented copy number aberrations at (A) diagnosis and (B) relapse. Processed log ratio data and B-allele frequency data for both (C) diagnostic and (D) relapse samples.

This case of relapsed BL was noteworthy for its extremely quiet genome at both diagnosis and relapse. There were no copy number aberrations present at diagnosis

that were maintained at relapse. The fact that both abnormalities detected by array at relapse were subtle and also did not appear clonal suggests either that the sample has a low tumour content compared to the diagnostic sample, or that the copy number aberrations were subclonal. This case further supports the need for *TP53* aberration in disease progression with a G245S mutation being the only abnormality screened for that was detected at both timepoints. The pattern of evolution resembles the early diverging type as the diagnostic and relapse profiles were so different. Exome sequencing analysis could only be performed on the diagnostic sample as the FFPE DNA quality was insufficient. If this sample could be sequenced then more complex tracking of the resistant subclones could be performed by looking at the differences in VAF of mutations or employing more sophisticated tools.

7.3.8 Sporadic Burkitt lymphoma pair 7 – 4/136 and 10/136

Patient 4/136 was a 4-year-old male diagnosed with BL. The stage II tumour was abdominal, and the patient had no bone marrow or CNS involvement of the disease. The primary tumour was treated on FAB/LMB-based protocols in group B and completed CYM1 in remission. Three months later the patient relapsed, followed by another relapse four months after that. The patient died two months after the second documented relapse. Tumour material was available for the diagnostic sample and the two relapses, but the fragment size of the DNA from the second relapse was insufficient for array analysis or sequencing. The was also insufficient DNA remaining from the diagnostic sample for sequencing. The diagnostic sample harboured a gain of chromosome 7 and 17q CNN-LOH that were both maintained at relapse and deletion on 16p that was not present at relapse. 1q gain and 8p telomeric loss were acquired at relapse and were not detected at diagnosis (Figure 7.20).

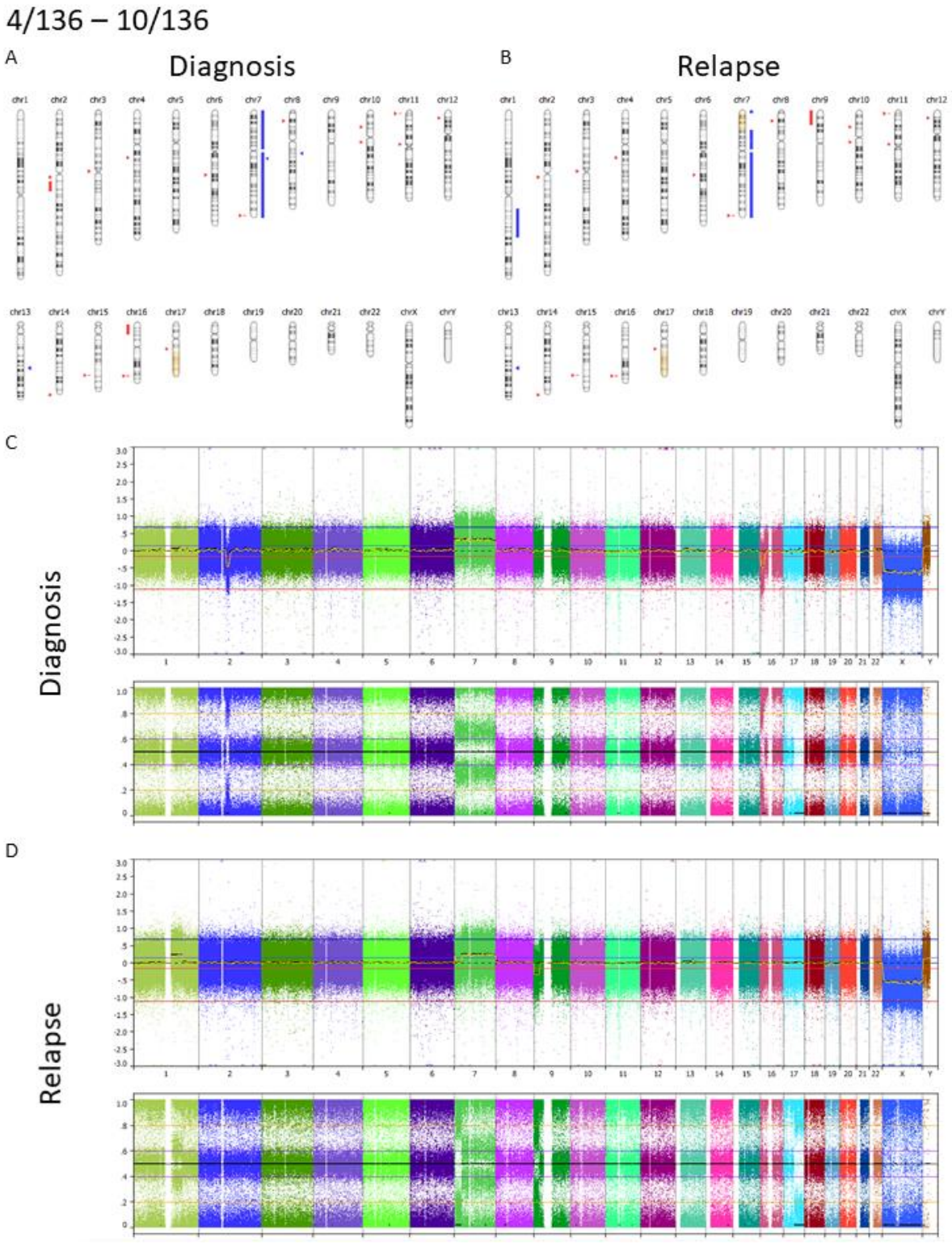


Figure 7.20 - Copy number profile of sporadic BL patient 4/136 and 10/136 at diagnosis and relapse. Summary of segmented copy number aberrations at (A) diagnosis and (B) relapse. Processed log ratio data and B-allele frequency data for both (C) diagnostic and (D) relapse samples.

A region of copy number gain on chromosome arm 1q typical of the abnormalities identified by GISTIC2.0 was present in the relapse sample but not the diagnostic sample, even at low level. This is surprising as 1q is a common abnormality in BL but

not prognostic and is likely a passenger abnormality. A 2q12.1-2q14.1 deletion was also present at diagnosis but not relapse. The BAF tracks in both cases suggest that the aberrations were present in the entire sample. The whole of chromosome 7 was gained at diagnosis and this was maintained at relapse. However, an additional event occurred between the diagnostic sample being taken and the first relapse biopsy where one copy of the region between 7p22.1 and 7p15.1 was deleted, making it copy number normal (Figure 7.21). This was confirmed by CNN-LOH arising in the region that was originally deleted. This arose due to the non-duplicated copy of the region being deleted, leaving behind the original and duplicated version. This provides some evidence of a model of tumour evolution with an ancestral clone that had already acquired the extra copy of chromosome 7 at diagnosis.

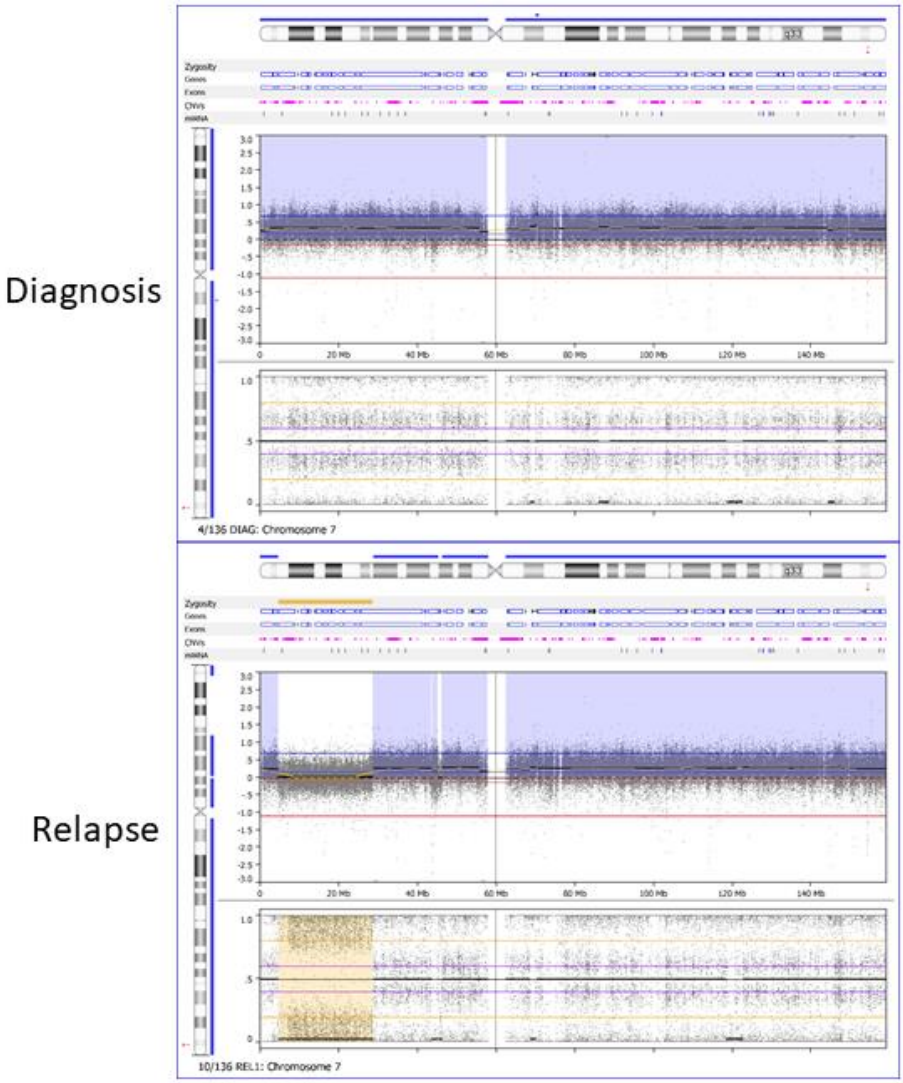


Figure 7.21 - Log ratio and B-allele frequency data for chromosome 7 depicting a copy number gain present at diagnosis in the relapse. Subsequent events led to the loss of one copy of 7p22.1-7p15.1 at relapse with associated CNN-LOH in the region.

A 9p deletion from the telomere to 9p21.3 was found at relapse in which the breakpoint was adjacent to *CDKN2A* and one copy of the gene was lost. The allelic frequency track showed that the deletion was not fully clonal and that a proportion of the tumour sample still retained heterozygosity at the site. A deletion was seen on 16p at diagnosis but was not maintained at relapse, despite being present in almost the entire tumour population. 17p was normal in both samples while 17q CNN-LOH was maintained from diagnosis to relapse with the same breakpoint at 17q21.31.

7.3.9 DLBCL paired analysis

DLBCL is most comprehensively studied in an adult setting where most diagnoses occur, while paediatric relapsed/refractory DLBCL is poorly understood. In a study of 20 DLBCL relapse pairs 65% of cases were found to be late divergent or linear in nature (Juskevicius *et al.*, 2016), similar to the BL relapses observed in this investigation. 35% of DLBCL cases in that study had early divergent evolutionary patterns where the relapsed tumour bore little resemblance to the original diagnostic sample, and three cases (15%) had entirely unrelated relapses. Two DLBCL relapse cases were available for analysis in this investigation and both had sequential biopsies at both diagnosis and relapse. As the two relapse/refractory DLBCL pairs underwent disease progression early (2.3 and 4.3 months respectively) it was not expected that either would be clonally unrelated recurrences.

7.3.10 DLBCL pair 1 – 16/547

Patient 16/547 was diagnosed at age 16 with DLBCL localised to the intestines. Case notes suggested that the patient had an underlying common variable immunodeficiency, which was likely an important driver for lymphomagenesis in this patient. The stage III DLBCL was negative for MYC, BCL2 and BCL6 rearrangements. The patient was treated at frontline on an adult protocol of R-CHOP alongside with rituximab at relapse. There was no recorded date of remission, but

clinical data suggested the patient relapsed rather than developed primary refractory disease. The patient died from disease nine months after relapse. The diagnostic and relapse samples did not resemble each other, with 1p CNN-LOH, 17p CNN-LOH and a deletion on chromosome 10 maintained from diagnosis to relapse (Figure 7.22).

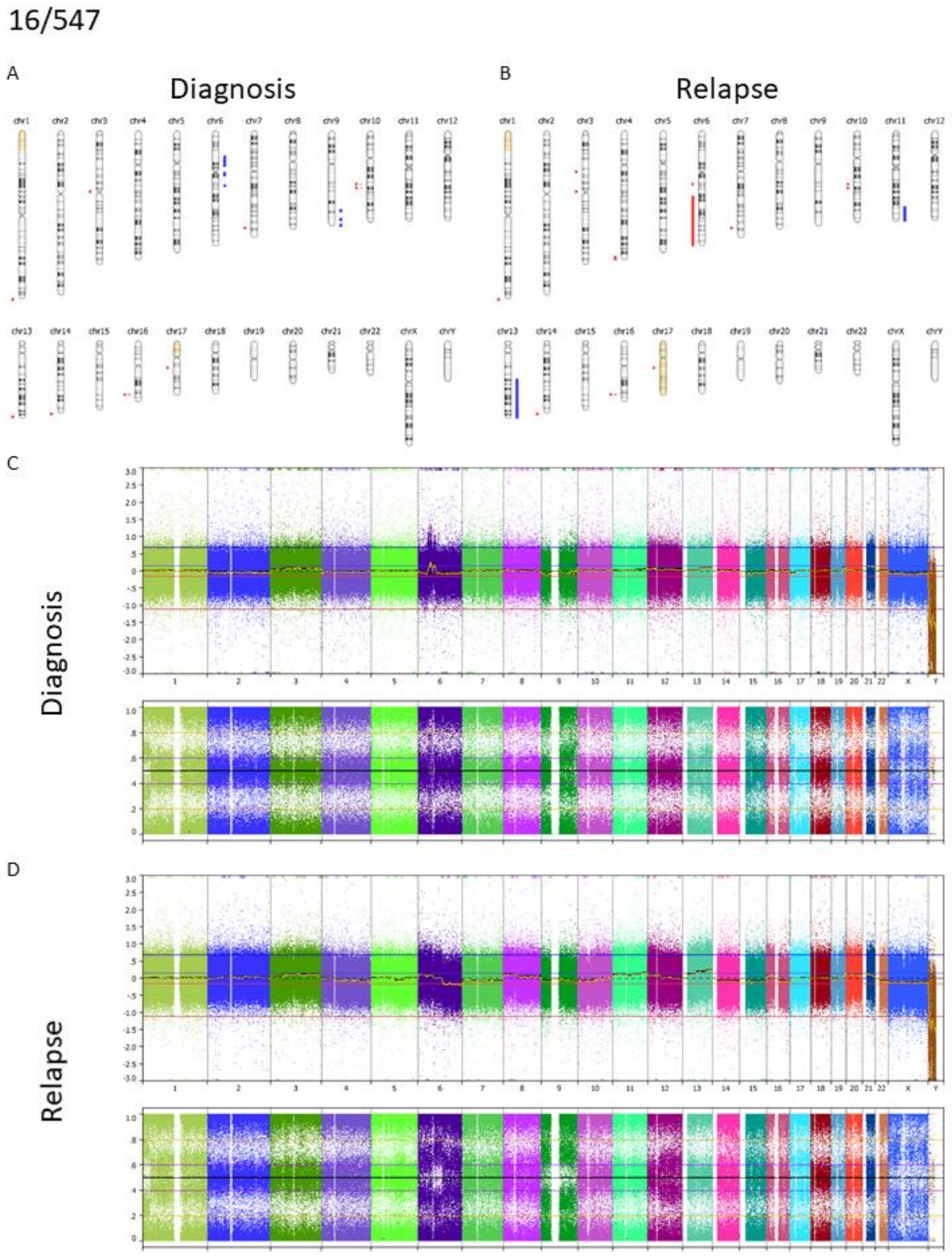


Figure 7.22 - Copy number profile of DLBCL patient 16/547 at diagnosis and relapse. Summary of segmented copy number aberrations at (A) diagnosis and (B) relapse.

Processed log ratio data and B-allele frequency data for both (C) diagnostic and (D) relapse samples.

A small region of CNN-LOH at the 1p telomere from 1p36.33 to 1p35.3 was observed at diagnosis and maintained at relapse. On top of this at diagnosis it was clear that a small subclonal population had CNN-LOH of the whole chromosome, which was maintained within a slightly higher proportion of cells at relapse but still not the clonal majority. The frequency of 1p CNN-LOH within the relapse simple was higher than at diagnosis. A complex pattern of gain around the centromere of chromosome 6 was seen at diagnosis, involving high copy number amplification between 6p21.2 and 6q12. An additional region of amplification at 6q14.1 was also seen at diagnosis involving the *TTK* and *BCKDHB* genes. None of the amplified regions seen at diagnosis were amplified at relapse. Instead the majority of 6q was deleted from 6q16.1 to the telomere, which equally was not present at diagnosis.

Chromosome 9 had aberrations at both diagnosis and relapse that were not consistent between timepoints. The diagnostic sample harboured low level CNN-LOH on 9q with three well defined copy number gains on the arm at 9q32-9q33.1, 9q33.3-9q34.11 and 9q34.3 respectively. The amplification had consistent log ratio values of 0.18, 0.15 and 0.17 respectively, suggesting that they coexisted in the same cell population, possibly as a result of a chromothripsis-like event. These aberrations were not present at relapse, but the whole of chromosome 9 had acquired CNN-LOH in a subpopulation, not just 9q. The proportion of the tumour with CNN-LOH was higher than at diagnosis, alluded to by the BAF track data. Chromosome 13q was gained at relapse, but there is some suggestion that the aberration was present at diagnosis in a very small population and went undetected by the segmentation algorithm. The whole chromosome was gained in three segments with the highest point of gain at relapse occurring from 13q31.3 to the 13q telomere.

17p CNN-LOH was visible in a small population at diagnosis. This was seen at relapse also, but the CNN-LOH affected both arms of the chromosome at this timepoint. Interestingly *TP53* was not mutated at both diagnosis and relapse, which may suggest an alternative gene of interest was mutated at diagnosis to confer the survival advantage to cells that had acquired 17p CNN-LOH. CNN-LOH on 17p at diagnosis developed into CNN-LOH of the whole of chromosome 17. This may have

occurred as a secondary event within the subclone that harboured 17p CNN-LOH at diagnosis or may be the result of a distinct subpopulation arising.

The evolution of disease progression within this case appeared to be early divergent, where the genomic profile of the tumour at diagnosis did not resemble the sample at relapse.

7.3.11 DLBCL pair 2 – 5/894

Patient 5/894 was a 4-year-old female diagnosed with DLBCL. They were diagnosed with a stage III DLBCL of the abdomen and pelvis with no bone marrow or CNS involvement and IG-MYC rearrangement was confirmed by FISH. The patient was treated on FAB/LMB-based protocols in group B but was escalated to group C after a lack of complete response to COP1 treatment. The patient did not achieve remission following CYM1/CYVE2 therefore was moved to salvage therapy for refractory disease, comprised of RICE and BEAM protocols. This salvage treatment was successful, and the patient has gone three years since diagnosis without further progression. An additional copy of chromosome 12 was observed at diagnosis and relapse, while the relapse sample also acquired deletions of chromosomes 6 and 10 (Figure 7.23).

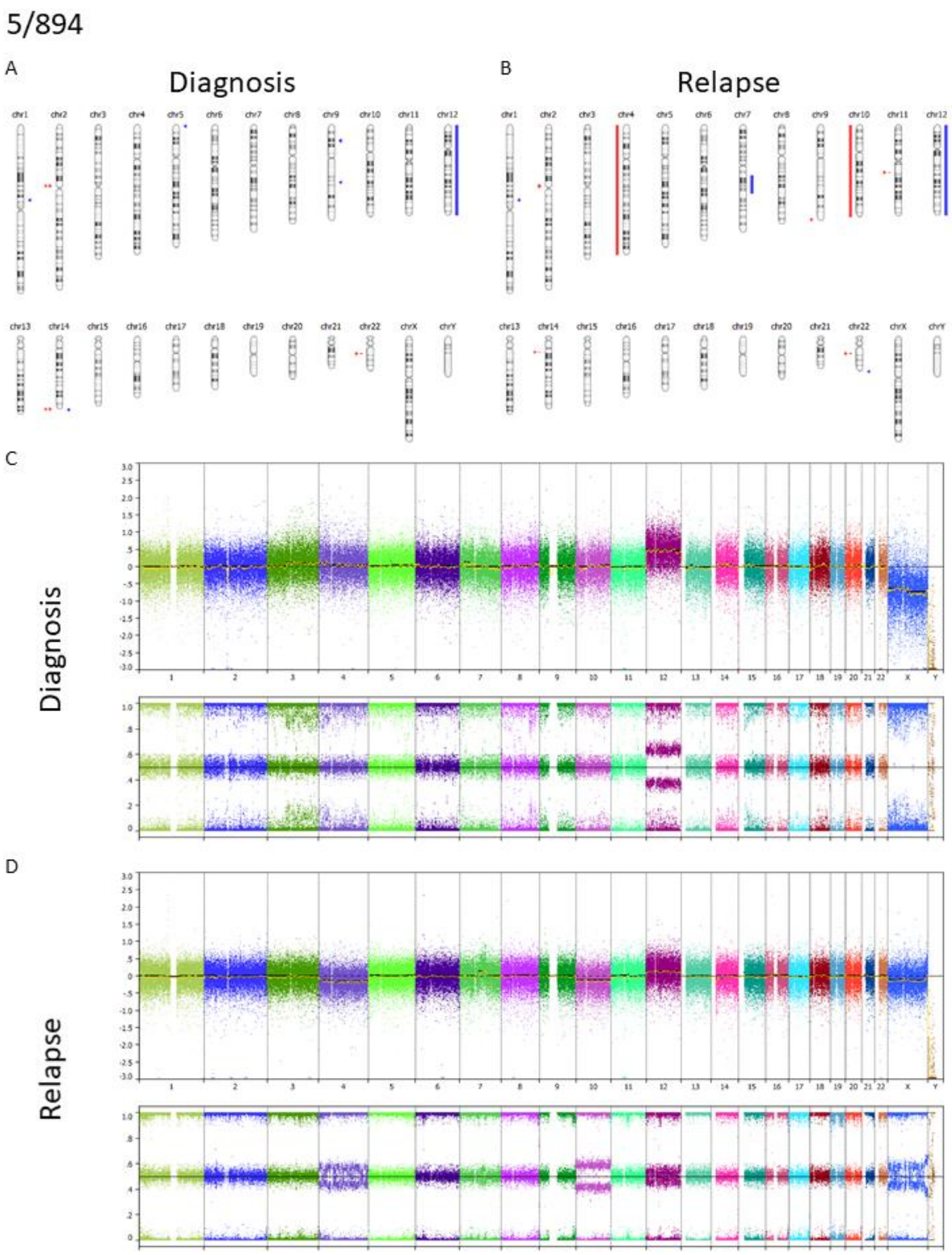


Figure 7.23 - Copy number profile of DLBCL patient 5/894 at diagnosis and relapse. Summary of segmented copy number aberrations at (A) diagnosis and (B) relapse. Processed log ratio data and B-allele frequency data for both (C) diagnostic and (D) relapse samples.

5/894 harboured a low-level copy number gain of chromosome 3 at diagnosis that was not detected by the TuScan segmentation algorithm with a low log ratio of 0.11.

This abnormality spanned from 3p12.3 to the entirety of 3q. The relapse sample showed no evidence of this event, inferring that the population involved did not survive treatment. Conversely, chromosome 4 was deleted in a subclone of the relapse tumour but not at diagnosis. 7q11.23 to 7q22.1 was amplified at relapse but not diagnosis and This region has been previously associated with outcome in B-NHL in cytogenetic studies (Poirel *et al.*, 2009; Nelson *et al.*, 2010). The miniscule deviations in log ratio from the origin in these samples suggests a low tumour cell content within the samples. The exception to this finding was gain of chromosome 12 which was shared by both diagnostic and relapse samples. At diagnosis the log ratio of chromosome 12 was significantly higher than at relapse and was clearly present in significantly more of the sample than gain on chromosome 3, suggesting that chromosome 12 gain was common to a large portion of the tumour and that tumour content was high in this sample. It was not possible to comment on the tumour content of the relapse sample from array data alone as there were no aberrations with high deviations in log ratio to determine the approximate log ratio of a clonal abnormality.

A small copy number gain at 9p21.3 was seen in the diagnostic biopsy but was not maintained. The telomeric breakpoint was adjacent to the *CDKN2A* gene and eight genes were within the segment. A subclonal deletion of chromosome 10 was seen at relapse but not diagnosis, meaning that gain of chromosome 12 was the only aberration seen in both samples.

7.3.12 Endemic Burkitt Lymphoma paired analysis

Paired material from diagnostic and relapse biopsies was available for two patients diagnosed at Queen Elizabeth Central Hospital in Blantyre, Malawi. The time at which relapse was identified in the endemic cohort may have been later than in the UK and the disease could have progressed further as the patient would not have received regular scans to screen for recurrent disease. Selection pressure on the tumour cells exerted by therapy was different to that of the B-NHL cases seen in the UK, so the pattern of clonal evolution cannot be directly compared. However, common factors may be identified between endemic and sporadic BL cases which are shown to drive disease progression. Other co-morbidities including HIV infection,

EBV infection, malarial infection, malnutrition and other confounding factors were more common in the endemic cohort, further separating the clinical situation between analyses.

7.3.13 Endemic Burkitt lymphoma pair 1 - B266/B300

Patient B266/B300 was a 3-year-old male diagnosed with a MYC-positive Burkitt-like lymphoma. IHC performed at Newcastle University retrospectively confirmed that the patient likely had a high-grade B cell lymphoma and not BL. The patient received one course of treatment out of four, consisting of vincristine, cyclophosphamide, methotrexate, hydrocortisone and doxorubicin (Molyneux *et al.*, 2017). The patient then absconded for two months which is not uncommon in this clinical setting. After two months the patient returned to the clinic with active disease and died shortly after. As the patient did not complete frontline therapy the disease cannot be considered to have relapsed. It is conceivable that the malignancy would have been successfully treated if frontline therapy had been completed. However, the case was informative with regard to the development of the tumour after a small amount of treatment. The patient also represents a significant clinical challenge in the Malawi setting; patient abscondment during treatment.

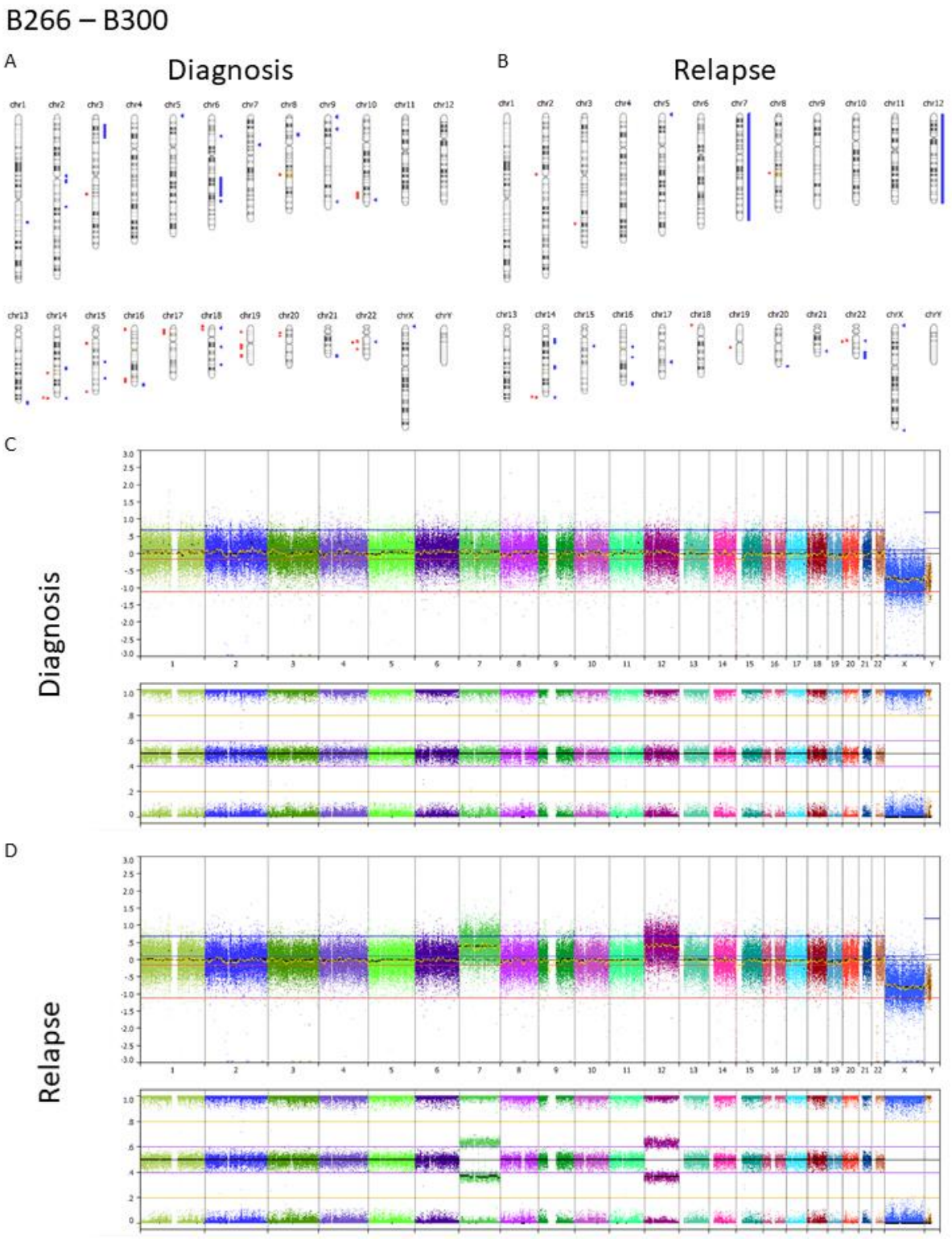


Figure 7.24 - Copy number profile of endemic HGBL patient B266/B300 at diagnosis and relapse. Summary of segmented copy number aberrations at (A) diagnosis and (B) relapse. Processed log ratio data and B-allele frequency data for both (C) diagnostic and (D) relapse samples.

The copy number profile of the diagnostic sample was almost completely normal with no large abnormalities (Figure 7.24A-D). There was a baseline level of noise that was

higher than in the CCLG tumour samples which was most probably a result of the different sample processing and storage technique at the QECH in Malawi. There were 177 small regions of copy number gain that were not supported by changes in the BAF track, suggesting that they were the result of natural noisy waviness in the data and not biological. The relapse sample for this patient, B300, harboured two abnormalities not present at diagnosis. The sample at relapse had a single extra copy of chromosomes 7 and 12 (Figure 7.24A-D). This is significant as both chromosomes were recurrently gained at relapse in the sporadic and endemic cohorts. No other aberrations were present in the relapse sample. It is clear, however, that chromosomes 7 and 12 play an important role in the development of non-Hodgkin's lymphoma.

7.3.14 Endemic Burkitt lymphoma pair 2 - B558/B594

Patient B558/B594 was a female patient who was 7 years old when diagnosed with typical endemic BL at QECH, Malawi. FISH confirmed 98.2% of cells had an IGH/MYC translocation. The patient received additional prednisolone as well as cyclophosphamide, vincristine, doxorubicin, methotrexate and hydrocortisone (Molyneux *et al.*, 2017). The patient relapsed four months after initial diagnosis where they received three courses of salvage therapy, before being palliated with morphine at home.

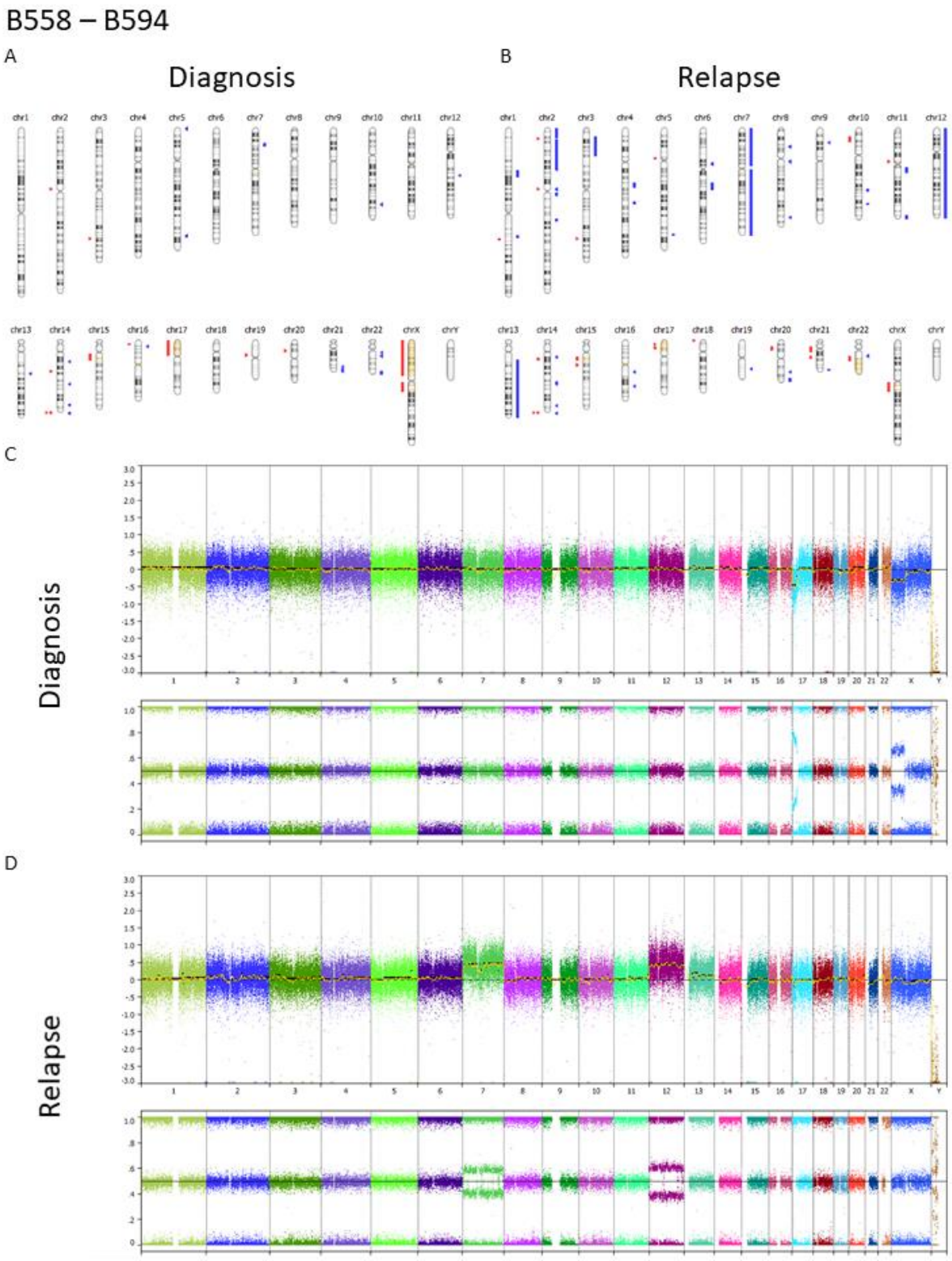


Figure 7.25 - Copy number profile of endemic BL patient B558/B594 at diagnosis and relapse. Summary of segmented copy number aberrations at (A) diagnosis and (B) relapse. Processed log ratio data and B-allele frequency data for both (C) diagnostic and (D) relapse samples.

The diagnostic sample harboured few abnormalities and had a quiet copy number profile (Figure 7.25-D). A copy number gain on chromosome 13 involving the *FOXO1* gene was observed but the log ratio was low (probe median 0.11) and the allelic ratio

did not support any change of copy number in the region. In the relapse sample there was a subclonal gain of 13q, but this did not align with the putative region seen at diagnosis. The gain of 13q did not appear clonal in the relapse sample as the log ratio was 0.13 and the BAF track was not strongly affected. There was also deletion at the centromeric end of 15q involving a 10.1Mb region from 15q11.1 to 15q13.1 that was maintained at relapse. At relapse there were gain of the whole of chromosomes 7 and 12, similarly to B266. The gains appeared to be clonal with high log ratios (chr7: 0.47, chr12: 0.46) and the BAF data showed a clear deviation.

17p was deleted at diagnosis involving the whole chromosome arm. The aberration was not seen across the entire sample but affected a large portion of cells with a log ratio of -0.45. Remarkably the relapse sample had 17p CNN-LOH without loss that was entirely clonal, meaning that two distinct subclones arose that harboured 17p aberrations involving *TP53*.

7.3.15 Recurrent abnormalities in progressive disease

Figure 7.26 compares the copy number profiles of diagnostic and relapse samples diagnosed with sporadic BL. Gain on 3q involving 3q29 was seen in three relapse samples but was absent from diagnostic samples. Given that 3q29 gain was a recurrent copy number abnormality and was identified as a prognostic marker at diagnosis in Chapter 4, the fact that it was acquired in three relapse samples strengthens evidence for its prognostic value in B-NHL and suggests that it confers an advantage to the tumour against frontline therapy.

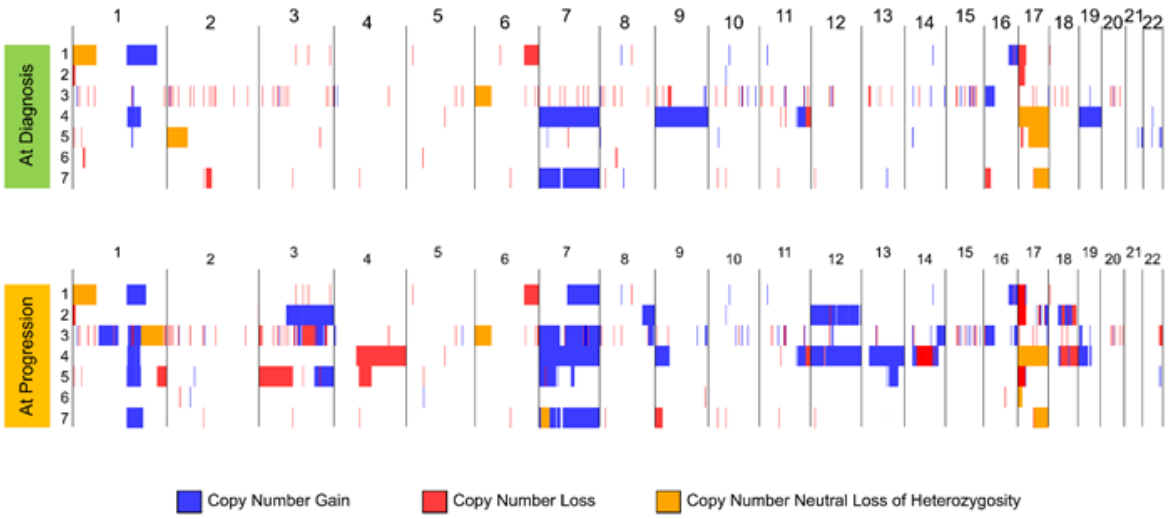


Figure 7.26 - Comparison of copy number profiles of paired diagnostic and relapse samples from seven paediatric sporadic Burkitt Lymphoma patients.

Chromosome 7 gain was seen in two cases at diagnosis and was maintained in both respective relapse samples. Three cases also acquired gains on chromosome 7 involving one 7p gain, one 7q gain and one gain of the whole chromosome. This strongly suggests a role for chromosome 7 in disease progression and resistance to chemotherapy. While a gain of chromosome 7 was not in itself prognostic in the cohort of 162 paediatric B-NHL diagnostic samples screened in chapter four, the abnormality was highlighted as a putative prognostic marker in the FAB/LMB96 trial (Poirel *et al.*, 2009). Chromosome 7 aberrations may arise frequently at relapse as passenger abnormalities, but do not confer a protective advantage to the lymphoma.

Three out of seven sBL relapse/refractory cases (42.9%) harboured a deletion of the *TUSC7* gene that arose at relapse (Figure 7.27). Deletion of this gene was seen in 19 diagnostic samples analysed in Chapter 4, primarily BL (12/19, 63.2%), but it is remarkable that the abnormality arose at relapse in 3/7 sBL relapse pairs. A recent whole genome sequencing study identified aberrations of the 3q13.31 locus involving *LSAMP* and *TUSC7* and determined they did not co-occur with *CCND3* mutations (Lopez *et al.*, 2019). The role of *TUSC7* aberration in relapse remains to be investigated fully.

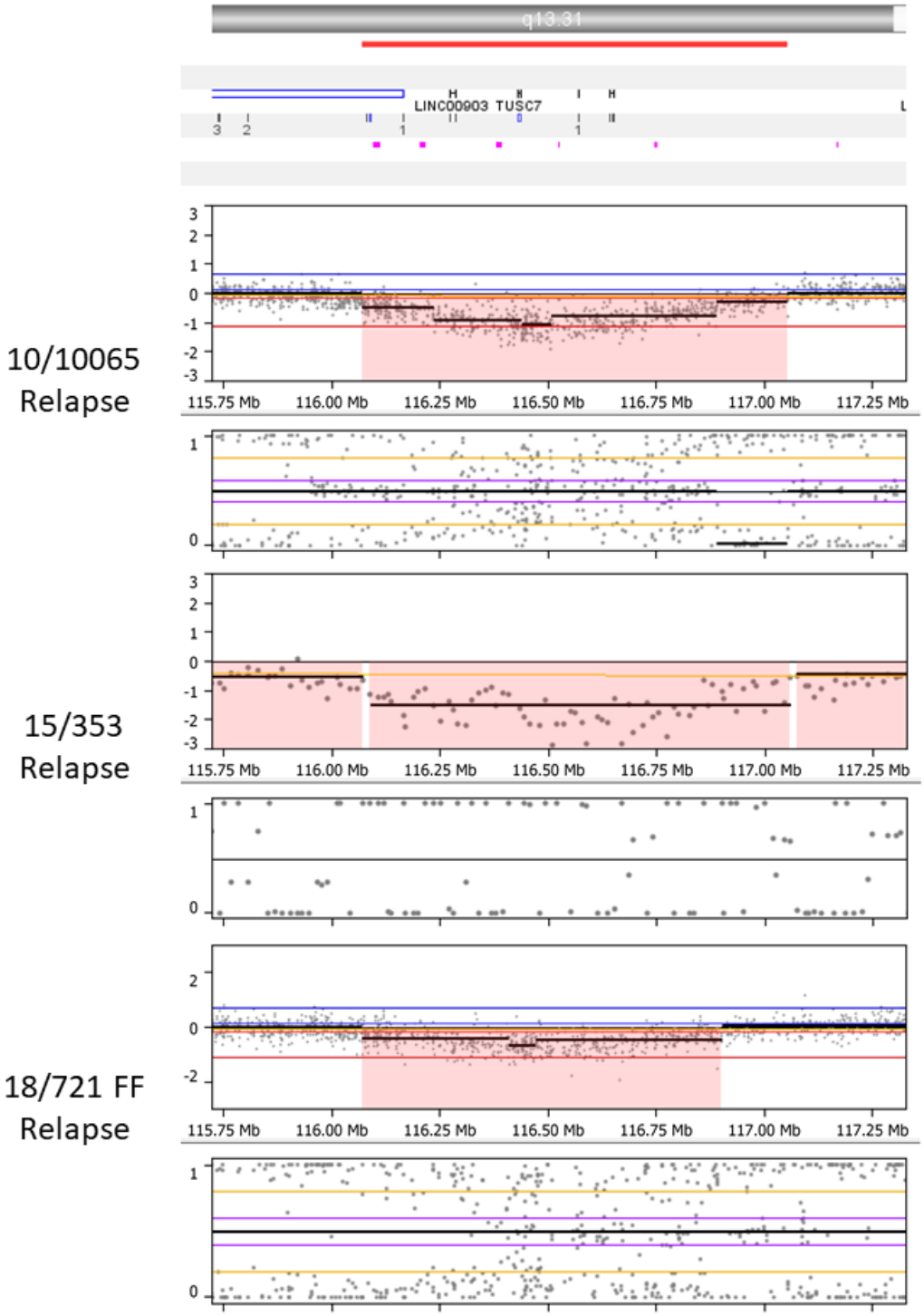


Figure 7.27 - Deletion on 3q involving TUSC7 was acquired in three relapse samples. Region shown spanned chr3:115723703-117327885.

Gain of 1q has long been known to be a common aberration in B-NHL, particularly BL (Lones *et al.*, 2004; Poirel *et al.*, 2009; Scholtysik *et al.*, 2010; Scholtysik *et al.*, 2015). Copy number gain of the region identified by GISTIC2.0 was seen in the diagnostic sample of two sBL pairs and maintained at relapse, with three additional cases with 1q gain events acquired at relapse. 1q gain has not been shown to be prognostic in large clinical trial cohorts (Poirel *et al.*, 2009) nor was it prognostic in

this paediatric diagnostic cohort of B-NHL cases and was determined to be a likely passenger abnormality. Investigation of 95 paediatric B-NHL cases with known *TP53* status showed that complex aberrations of 1q involving copy number gain were significantly associated with *TP53* status, which is a possible explanation for the increased incidence of 1q gain at relapse. *MDM4* amplification on 1q may confer a survival advantage to the tumour at during therapy, explaining the higher incidence of 1q gain at relapse as it may inhibit p53, but is also likely that 1q gain is a passenger abnormality, given the high frequency at which it was observed in the diagnostic cohort in Chapter 4.

17q CNN-LOH was present in three diagnostic samples of seven sBL pairs, however it was only maintained at relapse in two of these and no cases acquired 17q CNN-LOH. In the sBL case (case 18/721) where 17q CNN-LOH was lost at relapse, the pattern of evolution resembled the early-divergent/branching form that has been described previously in DLBCL (Juskevicius *et al.*, 2016). In this sample 17p CNN-LOH arose instead at relapse which may suggest that while both aberrations were prognostic in a diagnostic cohort, 17p CNN-LOH involving *TP53* biallelic inactivation conferred a stronger survival advantage to the tumour than 17q CNN-LOH.

7.3.16 TP53 Aberrations in Disease Progression

Work described in Chapter 5 of this thesis showed that *TP53* aberrations at diagnosis are significantly associated with a higher risk of disease progression and poor outcome in paediatric B-NHL. The exception to this observation was in endemic BL, where *TP53* abnormalities were not associated with a higher risk of relapse, but *TP53* was a prognostic marker of poorer overall survival. We hypothesise, therefore, that *TP53* abnormalities identified at diagnosis are maintained at relapse as they drive disease progression in a therapeutic setting. Where material was available, the *TP53* status of diagnostic and relapse pairs were investigated by copy number array and sequencing.

In patient 10/10065 the *TP53* status was stable throughout timepoints, with 17p loss and hemizygous G266E mutations with a VAF of 85% at diagnosis and 92% at

relapse (Figure 7.28C), meaning that both alleles of the *TP53* gene were aberrant at both diagnosis and relapse.

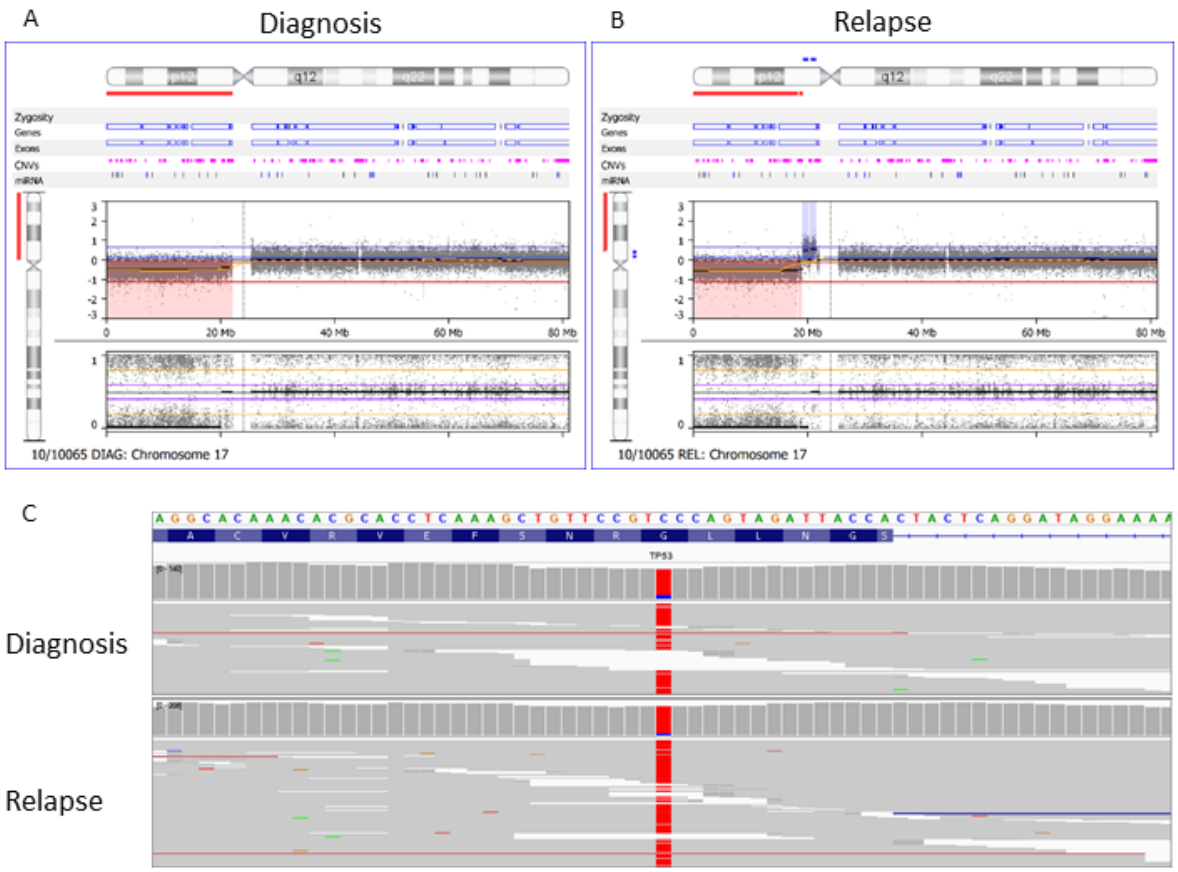


Figure 7.28 - Deletion of 17p was seen at both (A) diagnosis and (B) relapse in patient 10/10065, but a distinct pattern of copy number gain adjacent to the deleted region arose at relapse. (C) A G266E mutation in the TP53 *gene was identified at diagnosis and maintained at relapse.*

Similarly to 10/10065, both the diagnostic and relapse samples of 16/818 harboured a deletion of 17p and a mutation in on the remaining allele. The hemizygous V216G mutation in *TP53* identified at diagnosis was maintained at relapse and the VAF was slightly higher from 86% of reads to 92% (Figure 7.29C). The breakpoint of loss on 17p was different between samples with the diagnostic sample spanning just 17p13.3-17p11.2 and the relapse sample having loss of the whole arm all the way to the telomere (Figure 7.29).

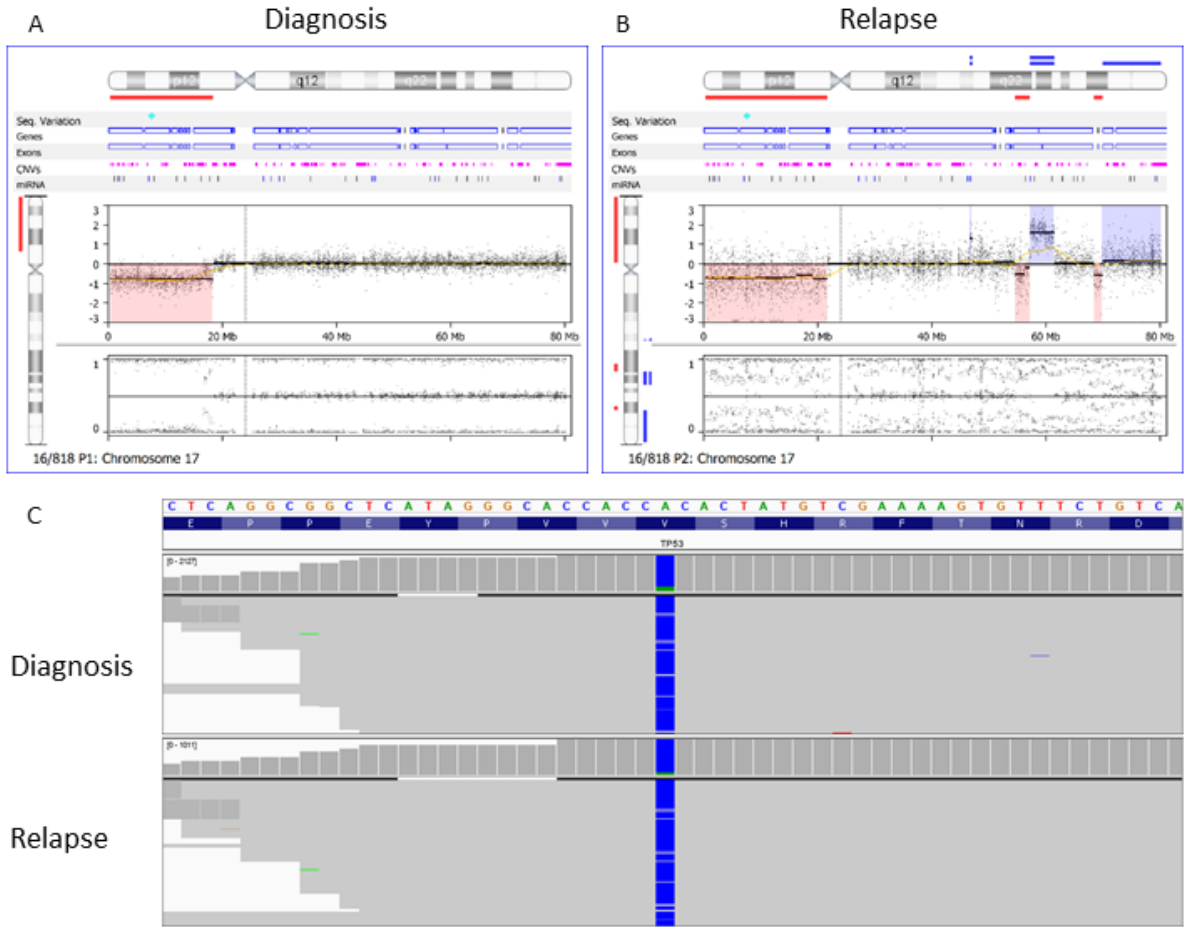


Figure 7.29 - Deletion of 17p was seen at both (A) diagnosis and (B) relapse in patient 16/818, but a distinct pattern of copy number gain and loss arose on 17q at relapse. (C) A V216G mutation in TP53 *was identified at diagnosis and was maintained at relapse.*

No copy number abnormalities involving chromosome 17 were seen in either sample from patient 15/353, but the diagnostic sample exhibited two *TP53* mutations at R248Q and R273C. The mutations were maintained at relapse, but the VAF increased significantly in one mutation. The R248Q mutation only increased from 44% to 49% suggesting that it was common to the majority of the tumour cells. In contrast the R273C mutation was only seen in 20% of reads at diagnosis and at relapse this population had expanded to 44%. Comparing this to the R248Q mutation which remained stable suggests that the R273C mutation was present in a subclonal population that already had a R248Q mutation in *TP53*. The two variants could not be shown to be present on different alleles, but this assumption was made based on the findings of Chapter 5. This is further evidence for *TP53* biallelic aberration being selected for by intensive chemotherapy, giving this subclone an advantage over those without *TP53* aberrations in the context of therapy.

Patient 16/391 harboured a R248Q mutation that we identified by whole exome sequencing of the diagnostic fresh frozen tumour sample and was confirmed to be present in the FFPE relapse sample by both Sanger sequencing and Haloplex targeted sequencing (Figure 7.30). This suggests that the resistant clone in this case harboured *TP53* biallelic aberrations.

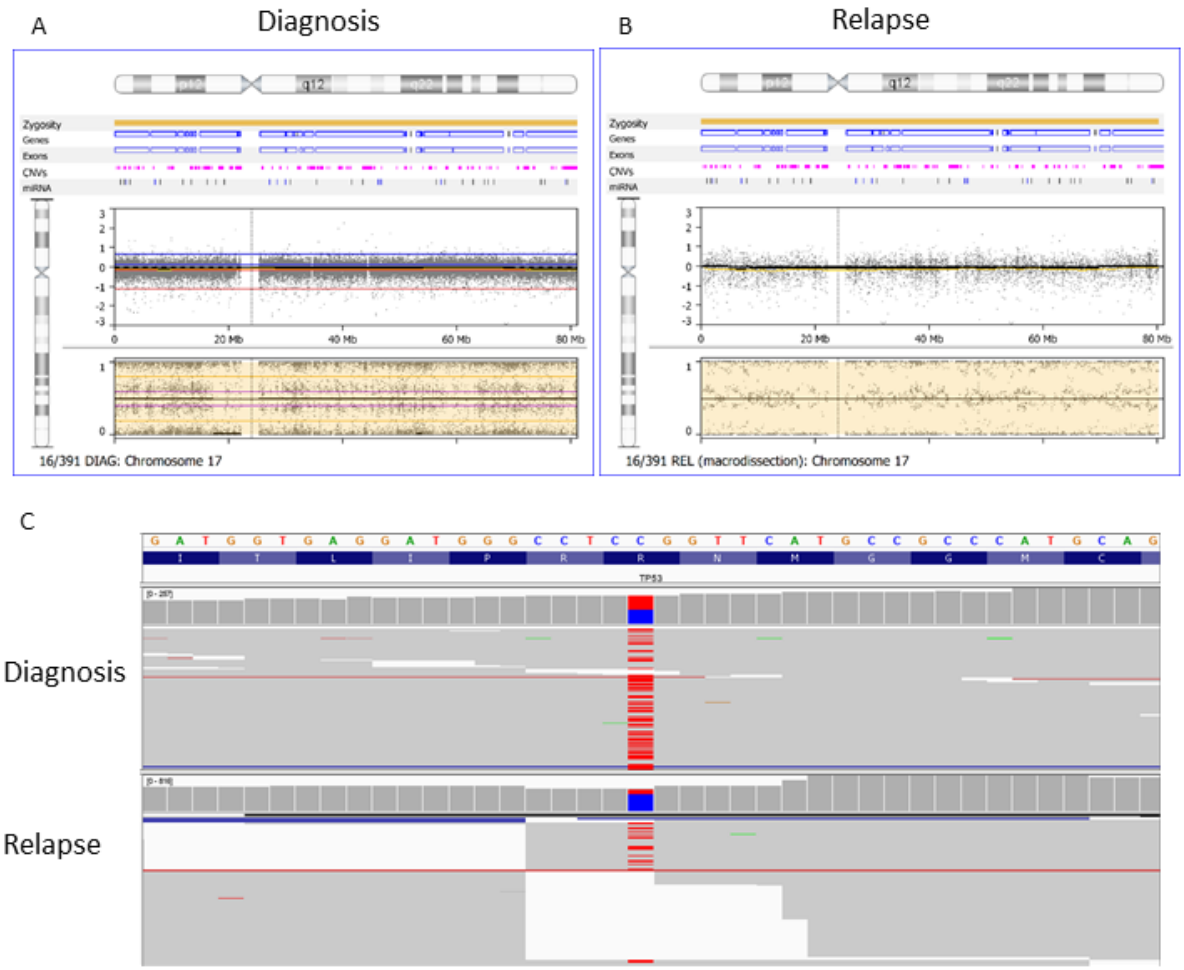


Figure 7.30 - Deletion of 17p was seen at both (A) diagnosis and (B) relapse in patient 16/391 but a distinct pattern of copy number gain and loss on 17q arose at relapse. (C) A V216G mutation in TP53 *identified at diagnosis was maintained at relapse.*

While a small deletion on 17p was seen at diagnosis in patient 18/721, it did not involve the *TP53* gene (Figure 7.31A). The diagnostic sample harboured CNN-LOH across the whole of the 17q arm with two distinct segments overlapping. Remarkably the relapse tumour samples both show a clonal 17p loss involving *TP53* that appeared to be clonal in both samples (Figure 7.31B-C) and exome sequencing of the FF sample revealed the presence of an R248W *TP53* mutation (Figure 7.31D).

Exome sequencing confirmed that there was no mutation in *TP53* at the point the diagnostic biopsy was taken, nor was there loss or CNN-LOH of the 17p arm (Figure 7.31D). This suggests that genomic evolution occurred between the point of diagnosis and relapse. The major clone at diagnosis was not maintained and receded during therapy.

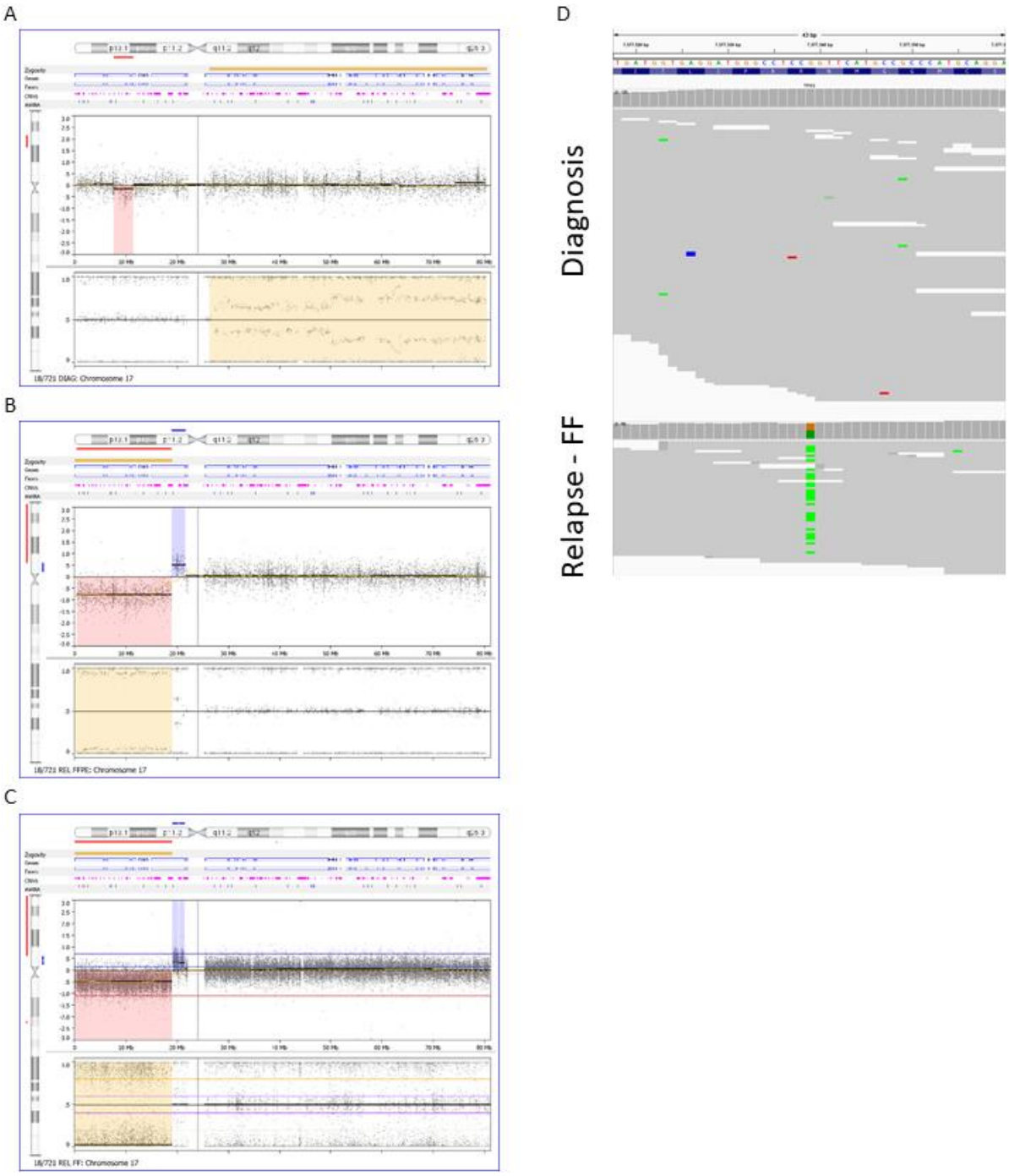


Figure 7.31 - Log ratio and B-allele frequency data for chromosome 17 in the (A) Diagnostic sample, (B) FFPE sample and (C) FF sample from 18/721. Both relapse sample harboured deletion of the 17p arm while the diagnostic sample showed a small deletion of 17p13.1-17p12 adjacent to but not involving TP53. The diagnostic sample also harboured 17q CNN-LOH which was not present in either relapse

sample. (D) Aligned whole exome sequencing showing the presence of an R248W mutation in TP53 that arose at relapse and was not detectable at diagnosis.

Between diagnosis and relapse the patient went from having two unaltered alleles of *TP53* to biallelic abrogation through two separate events. Given the fact that the 17p loss clonal line likely expanded from an undetectable subclonal population at diagnosis to a clonal effacement of the relapse tumour it is a sensible assumption to make that *TP53* was a vital driver of chemotherapy resistance in this case. The *TP53* mutation observed at relapse was not detected in a single read of the exome sequencing data at diagnosis.

The only abnormality that was acquired at relapse by patient 15/310 was 17p CNN-LOH in a small subpopulation of the relapse sample. There was no evidence of 17p aberration at diagnosis and the CNN-LOH was only present in a small subclone of the tumour at relapse (Figure 7.32). Exome sequencing analysis revealed that the patient had a somatic heterozygous G245S mutation in *TP53* at diagnosis which was found to be maintained at relapse by Sanger sequencing (Figure 7.32).

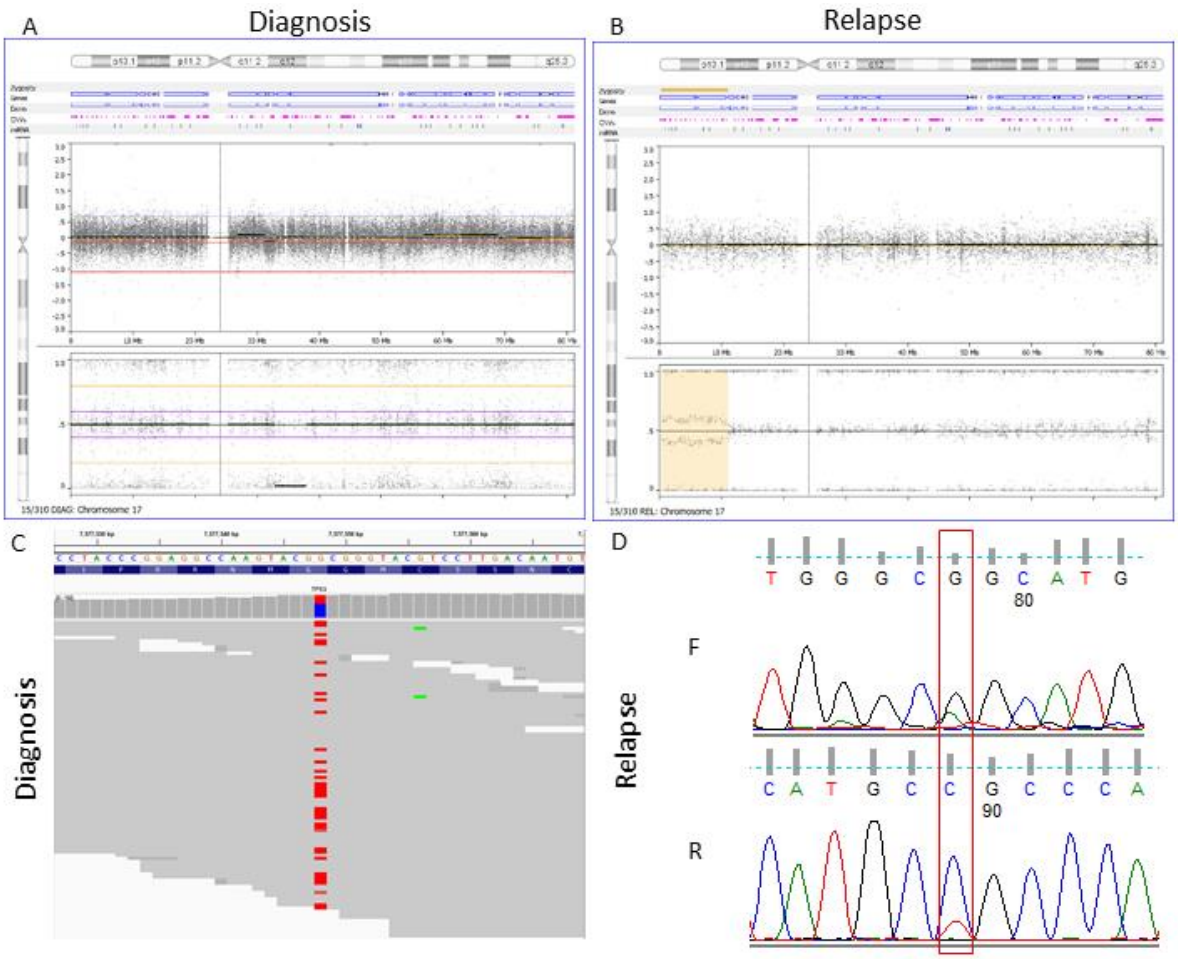


Figure 7.32 - CNN-LOH of 17p was not present at (A) diagnosis but was acquired at (B) relapse in case 15/310. A G245S mutation was present at (C) diagnosis in 39% of reads and confirmed to be present at (D) relapse by Sanger sequencing.F = Forward direction, R = Reverse direction.

Combined comparison of seven sBL relapse pairs highlights a strikingly recurrent element of disease progression: *TP53* abnormalities. All BL relapses that were screened had at least one *TP53* aberration at some point during their disease development and were always maintained as the lymphoma progressed. This further supports the suggestion that abrogation of p53 protects the cancer cell from FAB/LMB-based treatment and shapes the evolution of the tumour. Tumours that acquired *TP53* aberration at diagnosis followed a linear pattern of evolution with a relapse tumour that was closely related to the dominant diagnostic clone. With the exception of 4/136 all BL pairs either had biallelic aberration of *TP53* at diagnosis or acquired it at relapse, and in the exception, there was a *TP53* mutation at relapse even though it was not shown to affect both alleles.

In 16/547 there was no aberration involving *TP53* at diagnosis or relapse, which was a hallmark of BL relapsed pairs and a significant prognostic marker at diagnosis. It is possible that mutation in another gene on 17p is responsible for resistance to treatment in this case. 17p CNN-LOH was the only abnormality that was maintained from diagnosis through to relapse, although this may be coincidental as the context of 17p CNN-LOH was different at relapse with all of chromosome 17 affected and not just the p arm. Similarly, chromosome 17 of 5/894 was normal by array at both timepoints, with no CNN-LOH or deletion observed. A G187S mutation in *TP53* was identified at diagnosis by exome sequencing with a VAF of 39% and while this was maintained at relapse, the VAF dropped to 20%, suggesting that *TP53* was not a driver of disease progression and chemoresistance in this case.

Given that the *TP53* status of both 16/547 and 5/894 remained stationary between diagnosis and relapse gives credence to the hypothesis that *TP53* is important in tumour survival in BL but not DLBCL in the paediatric setting. This is particularly interesting as both malignancies are treated uniformly with FAB/LMB-based protocols in the UK, where these patients were sourced. It is possible that the strong pro-apoptotic signalling in BL caused by high MYC expression requires a stronger anti-apoptotic driver to overcome the stress of intensive chemotherapy, while in DLBCL alternative pathways give rise to disease progression independent of *TP53*. The ant-apoptotic pressure conferred by a heterozygous *TP53* mutation in this patient may provide a small survival advantage to the lymphoma as it was shown to have an IG-MYC translocation. However, it is clear that in the two paediatric DLBCL cases investigated that *TP53* is not as important in tumour survival as it is in the setting of paediatric BL.

Sanger sequencing of *TP53* performed previously in the lab showed that the diagnostic sample of endemic BL patient B558 had a heterozygous R248W mutation. Additional Sanger sequencing of the relapse sample confirmed that the R248W mutation in *TP53* was maintained at relapse and was homozygous. This suggested strongly that the mutation and deletion of 17p seen at diagnosis existed in distinct subclones and that mutation of *TP53* conferred a stronger survival advantage than hemizygous deletion, further intensified by a subsequent CNN-LOH event at relapse, leading to biallelic inactivation of the *TP53* gene. Frontline therapy to treat BL in Malawi is much less intensive than that which is given in the UK which means higher

relapse rates and lower overall survival (Molyneux *et al.*, 2012). The selection pressure conferred by treatment is therefore weaker than that in sBL relapse cases, reducing the advantage *TP53* aberration confers to the tumour.

7.4 Discussion

Comparing diagnostic and relapse samples in a non-paired manner allowed for identification of aberrations that were selected for at relapse, and that are likely candidates to drive disease progression and confer resistance to chemotherapy. It was clear through analysis of the seven sporadic BL relapse/refractory pairs that *TP53* aberration was the strongest factor associated with BL relapse, with particular emphasis on biallelic aberrations. Six out of seven pairs had biallelic aberrations at relapse making *TP53* abnormalities an important biomarker for screening. *TP53* status was also the most important factor in determining the evolutionary model that drove disease progression in sporadic BL. All four of the sporadic BL cases that had already established biallelic abnormalities involving *TP53* progressed in a late-diverging/linear evolutionary pattern. The two sBL patients that did not have biallelic aberrations at diagnosis but acquired them at relapse had early-diverging/branching clonal progression.

Copy number abnormalities identified at diagnosis in sBL were consistent with those at relapse, although additional events were seen in most cases with a more complex genome at relapse. No case lost a *TP53* aberration that was present at diagnosis; all were maintained at relapse, strongly supporting the idea that *TP53* is vital for BL survival. Interestingly, *TP53* status did not correlate with time to progression, as might be expected given the linear evolution pattern that biallelic *TP53* aberrations were associated with. Patient 16/391 took 6.1 months from diagnosis to relapse which was longer than either of the sporadic BL cases without biallelic *TP53* aberrations at diagnosis. Of these two cases, it is remarkable that both acquired additional *TP53* abnormalities, showing that even when the main clonal population of the tumour has monoallelic aberrations affecting the gene, further aberrations are still selected for in a treatment setting.

This is the first genomic study of disease progression across multiple timepoints in sporadic BL using copy number arrays. The only study to previously investigate the genomics of BL progression enlisted traditional cytogenetic techniques to compare sequential samples (Aukema *et al.*, 2015). The key findings of the study were the identification of a linear pattern of clonal evolution and an increase in genome complexity. Seven cases, including six paediatric and one adult case, were

compared between sequential biopsies. Five had diagnostic and relapse biopsies while two were compared at first and second relapse. An additional 18 BL cases with karyotype data from the Mitelman Database were analysed alongside their in-house cases, totalling 25 cases. Copy number abnormalities that have previously been reported in B-NHL were common at diagnosis and relapse, specifically gain of 1q, trisomy 7/7q and abnormalities involving 13q (Poirel *et al.*, 2009; Nelson *et al.*, 2010; Aukema *et al.*, 2015). Abnormalities on 11q, loss of 17p and loss of 19q were only seen at relapse. The findings of our study do not correlate directly with the genomic aberrations reported by Aukema *et al.* (Aukema *et al.*, 2015) as we identified two instances of 17p deletion and one case with 11q aberration at diagnosis. We do however report one instance of 17p deletion being acquired at relapse. We also report more complex genomes in sequential biopsies which may be driven by the increased incidence of biallelic *TP53* aberrations in sBL relapses, which we showed to be strongly associated with genomic instability in Chapter 5.

Despite clear evidence for a key role in sporadic BL progression, *TP53* status was stable in both DLBCL pairs. Both harboured one aberration that encompassed the *TP53* gene with 16/547 exhibiting 17p CNN-LOH and 5/894 having a somatic heterozygous G187S mutation. Previous studies into the clonal evolution of DLBCL relapse have reported an enrichment for *TP53* and *CDKN2A* aberrations at relapse (Juskevicius *et al.*, 2017), but this was not seen in either of our cases. No abnormalities could be shown to be significantly associated with relapse in DLBCL which highlights that a larger patient cohort is required.

In endemic B-NHL both patients acquired additional copies of chromosomes 7 and 12 at relapse with no evidence of the abnormalities at diagnosis. Previous studies in B-NHL into chromosome 7 gain, as well as analysis as part of chapter 4, have suggested that a minimal common region at 7q22.1 involves a gene of interest in B-NHL, most likely *MIR106B*, a paralogue of *MIR17HG*. p53 is a direct target of *MIR106B*, and inhibition of the constituents of the mir106b~25 miRNA cluster has been shown to increase endogenous p53 expression and causes cell cycle arrest and apoptosis (Kumar *et al.*, 2011; Mehlich *et al.*, 2018). Similarly, upregulation of *MDM2* on chromosome 12 at 12q15 may drive p53 inhibition via increased ubiquitination and degradation. More samples with matched diagnostic and relapse/refractory are required to comprehensively investigate disease evolution in

endemic BL. While *TP53* status for B266/B300 was unknown, the gene was sequenced in both samples from B558/B594. This patient had the most complex pattern of *TP53* aberration within the cohort, as the diagnostic sample had undergone a subclonal 17p loss as well as harbouring an R248W mutation in the *TP53* gene. At relapse the tumour acquired a 17p CNN-LOH, yet 17p deletion was still detected. Allelic frequency data from the copy number array data showed that the 17p CNN-LOH event was present in the vast majority of tumour clones, while 17p deletion remained in a subclonal population. It is possible that the subclonal population with a *TP53* mutation at diagnosis did not coexist with 17p deletions in the sample and acquired CNN-LOH at relapse. Without next generation sequencing data and the ability to compare allelic frequencies of *TP53* variants across samples, it is not possible to ascertain the pattern of development.

In this study we showed that disease progression in paediatric B-NHL follows one of two models previously reported and characterised in the literature (Jiang *et al.*, 2014; Juskevicius *et al.*, 2016; Juskevicius *et al.*, 2017). Remarkably, biallelic *TP53* aberrations were a strong determinant of the pattern of clonal evolution of relapse in sBL, suggesting that *TP53* aberrations are beneficial to the tumour survival under treatment. In DLBCL *TP53* abnormalities were present in both cases but were stable throughout disease progression. There was also evidence of *TP53* aberrations being involved in eBL relapse, with additional *TP53* abnormalities arising at relapse. In combination with the findings reported in Chapters 5 and 6, it clear that *TP53* is an important gene in paediatric B-NHL that drives disease progression in BL.

Chapter 8: General Discussion

Disease progression and treatment resistance represent the most significant clinical challenges in treating paediatric aggressive B-NHL (Kim *et al.*, 2014). Event-free survival in paediatrics is over 90% but the outcome for those that undergo disease progression is dire, with salvage rates as low as 17% (Fujita *et al.*, 2008; Poirel *et al.*, 2009; Anoop *et al.*, 2012). It is not currently possible to reliably predict which patients will be successfully treated on frontline therapy and who will undergo relapse/refractory disease, so prognostic markers are urgently needed that can be utilised in the clinic. Prognostic markers have been successfully used in the clinic in other cancers to stratify patients, for example in breast cancer HER2 and ER expression are both used clinically to predict poor outcome in patients (Mass *et al.*, 2005). This has been particularly successful in paediatric B-ALL, in which patients are stratified based on cytogenetic subgroups that are significantly associated with outcome (Rand *et al.*, 2011; Moorman, 2016).

Previous studies have attempted to address this requirement. A study of karyotyping data from 238 patients of the FAB/LMB-96 trial reported that additional copies of chromosome arm 7q and deletion of 13q were prognostic markers of disease progression and poor outcome (Poirel *et al.*, 2009). This represents the largest investigation of genomic markers of disease progression to date. A later study of 28 BL patients using copy number arrays investigated genomic aberrations associated with outcome. This was the first study of prognostic markers using high resolution arrays in aggressive B-NHL (Schiffman *et al.*, 2011). The study did not observe recurrent 13q loss or additional copies of 7q, instead reporting that *MIR17HG* amplification at 13q31.3 was associated with relapse. The observation was based on the finding that 2/3 relapse cases within the cohort had 13q31.3 amplification at diagnosis. However, no statistical analysis was performed due to small patient numbers. As with other genomic prognostic markers in B-NHL, this has not been translated to the clinic. The study also highlights the difficulty of identifying prognostic markers in small patient cohorts. No genomic studies involving meaningful cohort sizes have yet investigated outcome in paediatric B-NHL, with the majority of studies performed largely without clinical annotation or survival analysis, precluding the identification of useful prognostic markers.

Analysis performed in this thesis aimed to improve our ability to detect patients who were at risk of undergoing relapse/refractory disease at diagnosis. We hypothesised that copy number abnormalities detected by copy number arrays would be associated with disease progression and poor outcome. By combining our UK-based cohort of clinically-annotated patient samples with data from the literature the power of our investigation could be increased to create the largest clinically annotated dataset of high-resolution copy number data in paediatric B-NHL.

The pilot project described in Chapter 3 demonstrated that copy number data identified from the literature could integrated with in-house datasets to perform analysis on paediatric and adult B-NHL. The literature search highlighted the abundance of adult DLBCL copy number data, which is to be expected given the demographics of the disease (*HMRN - Incidence*, 2019). Conversely, paediatric cases were less common in the literature, with only 52 cases with copy number data identified. We expanded our cohort further by processing as many cases from the CCLG biobank as possible, expanding the patient cohort from 25 cases to 110 for analysis in Chapter 4. It was also clear that a systematic pathology review of patient material was required to validate the working diagnoses provided by the treating centre of treatment. Due to the fact that treatment of paediatric B-NHL in the UK is uniform, regardless of histopathological subtype, many patients in the CCLG cohort had incomplete diagnoses or were diagnosed using earlier editions of the WHO Classification (Swerdlow *et al.*, 2016). This initial pilot study highlighted significant data analysis hurdles that needed to be overcome, as well as issues regarding patient diagnosis and pathology review that informed and improved the later stages of the project.

Comprehensive analysis of copy number data from 162 paediatric B-NHL samples identified three independent prognostic markers: 3q29 copy number gain, 17q copy number neutral loss of heterozygosity (CNN-LOH) and 17p CNN-LOH. 3q29 copy number gain had been reported in BL previously (Scholtysik *et al.*, 2010) but was not shown to be prognostic. Copy number gains observed on 3q in the Scholtysik *et al.* study peaked at 3q27.2 and so were assumed to be associated with *BCL6* amplifications. However, in Chapter 4 we narrowed down the minimal common region of copy number gain to 3q29, and only 13/17 cases with this abnormality also harbouring an additional copy of *BCL6*. This suggests a different driver gene in the

region is responsible for increasing the risk of disease progression. Whole-exome sequencing data of a cohort of 39 B-NHL cases with paired tumour and normal sequencing data was analysed for mutations within 3q29. These data showed that somatic mutations involving *PAK2* and *OPA1* were recurrent in a cohort of 39 B-NHL cases with paired tumour and normal sequencing data, both of which were in the region of copy number gain in 17/17 cases with 3q29 gain. A study into *PAK2* in haematological malignancies reported that expression was highest in high-risk patients in BL and DLBCL (Edlinger *et al.*, 2017). Given this finding, the role of *PAK2* should be further investigated in paediatric B-NHL.

17q CNN-LOH was observed in 20/162 cases screened in Chapter 4. Two distinct focal regions were identified spanning 17q22-17q24.1 and 17q11.2-17q21.32. Investigating variants across 17q identified 57 recurrently mutated genes. *RNF213* was most frequently mutated but was not covered by a region identified by GISTIC2.0. *GNA13* was identified as a candidate gene within the region as mutations in *GNA13* are reported to be recurrent in both BL and GCB-like DLBCL (Love *et al.*, 2012; Schmitz *et al.*, 2012). 16/73 cases screened harboured a mutation in the gene, and 4/6 cases with CNN-LOH spanning the gene harboured mutations. Despite this association between 17q CNN-LOH and *GNA13* mutation, mutations of the gene were not associated with disease progression or outcome, but it may suggest that mutations in *GNA13* may cooperate with other abnormalities in disease progression.

Chapter 4 was performed to provide a genome wide screen of copy number abnormalities in a cohort of paediatric B-NHL patients and to integrate this analysis with well-annotated clinical data to identify markers of poor prognosis. A major downside to this analysis was the availability of clinical data. Every effort was made early in the project to acquire time to follow up, relapse status and a key set of clinical parameters. In particular, relatively few centres could supply information regarding bone marrow and central nervous system involvement, as well as LDH levels with paired normal range data. These clinical factors are currently used to decide what treatment arm patients are registered to, and as a result should be included as cofactors in multivariate survival analysis. It would be beneficial to perform this analysis with more clinical information available. Focused analysis of whole exome sequencing data was performed as part of the work reported in this chapter to prioritise regions believed to be associated with significantly inferior outcomes in

paediatric B-NHL. However the study would benefit from an exome-wide analysis of the same data to identify prognostic genetic variants which are not associated with copy number abnormalities.

17p CNN-LOH involving the *TP53* gene was also found to be a significant prognostic marker associated with disease progression and poor outcome. In Chapter 5 I investigated the role of *TP53* aberrations in paediatric B-NHL to determine whether this explained the prognostic value of 17p CNN-LOH. Mutations were present in 46/95 cases screened (48.4%). Sequencing and copy number array data were integrated to determine the *TP53* status of all 95 patients in the analysis. Overall, 54.7% (52/95) of cases had at least one aberration involving the *TP53* gene, with half of those (26/95, 27.4%) harbouring abnormalities affecting both alleles.

Remarkably, *TP53* aberrations were associated with lower overall survival and a shorter time to progression. *TP53* mutations alone were also a significant prognostic marker. In fact, the *TP53* normal group of the cohort had such a favourable prognosis that only one case underwent disease progression, and in Chapter 7 this case was shown to acquire two distinct aberrations involving the gene at relapse. The role of *TP53* in B-NHL has been investigated previously but has not been discussed in the context of outcome. Leventaki *et al.* (2012) postulated that all BL patients must have abrogated *TP53* function as CDKN1A expression was negative in 30/30 cases that they screened. Further characterisation of the cohort is required to determine whether functional p53 is still expressed in the cases without *TP53* aberrations or whether regulators including *MDM2* and *MDM4* are involved. Regardless of this, we have identified a clear association between genetic abnormalities of the *TP53* gene and a subsequent poor outcome. In chronic lymphocytic leukaemia, *TP53* mutations are routinely screened for at diagnosis and are used to direct treatment decisions (Campo *et al.*, 2018). This could easily be implemented in paediatric B-NHL, as the facilities already exist in hospitals across Europe (Malcikova *et al.*, 2018). *TP53* abnormalities are not targetable directly, but in chronic lymphocytic leukaemia (CLL) these cases respond well to PI3K signalling inhibitors (Idelalisib) and BCL2 inhibitors (Venetoclax) (Brown *et al.*, 2014; Stilgenbauer *et al.*, 2016). A clinical trial investigating the efficacy of Venetoclax in combination with PI3K inhibitor Copanlisib in relapsed/refractory B-NHL is currently recruiting in Switzerland, the results of which may provide supporting data for *TP53* aberrant lymphoma being a targetable

subgroup (ClinicalTrials.gov Identifier: NCT03886649). An important consideration with the potential use of these drugs on *TP53* aberrant BL is that it is not clear whether CLL patients respond well to Idelalisib and Venetoclax due to the *TP53* aberrant nature of the malignancy, or whether *TP53* simply highlights a high-risk group, meaning that targeted therapy is administered earlier. It is also important to mention that Idelalisib has been shown to be ineffective in BL. The key finding from the analysis performed during this project is that *TP53* identified a subgroup of approximately 50% of patients with extremely low risk of relapse/refractory disease who might be candidates for treatment de-escalation. Treatment toxicity is a major problem in treating paediatric B-NHL (Cairo *et al.*, 2018) and identifying patients who do not require the full protocol will help to reduce toxic effects.

The increased risk of disease progression conferred by *TP53* aberrations in paediatric B-NHL may function through increased genomic instability. We observed frequent complex chromosomal aberrations in patients with biallelic abnormalities of *TP53* using copy number arrays. Aberrations in the *TP53* gene are associated with genomic instability in medulloblastoma (Rausch *et al.*, 2012) and has been shown to drive chromothripsis in these patients. *TP53* mutations have also been shown to drive genomic instability in ovarian cancer, which directly drives chemoresistance and disease progression (Zhang *et al.*, 2017). Complex chromosomal abnormalities affecting chromosome arms 1q, 11q and 13q were significantly more frequent in the *TP53* biallelic group, but none of these were independently associated with outcome. Future investigations into the mutational profile of *TP53* abnormal vs normal subgroups will allow us to characterise these patients and the complex aberrations they harbour more comprehensively.

The work presented in Chapter 5 could be further improved by the inclusion of additional clinical data in the same way as Chapter 4. Considerable gaps in bone marrow, CNS and LDH data hampers the survival analysis and translation to the clinic and the study would be improved by reanalysis with more clinical data available. Additionally the cohort could be improved by securing constitutional DNA from more patients so that mutations could be confirmed to be somatic. Analysis of complex chromosomal aberrations highlighted an important association between *TP53* status and these aberrations. However, it was not possible to properly characterise these complex genomic changes using copy number arrays alone. Copy

number arrays are not able to define breakpoints in complex aberrations, and are relatively low resolution. To account for this, a sequencing-based approach would be necessary. Whole genome sequencing would be ideal as many of the breakpoints would not be within exons, and long read sequencing would allow for accurate mapping of aberrations. Oxford Nanopore's MinION technology would be an effective tool for analysing these aberrations.

An important finding from survival analysis in paediatric B-NHL was the observation that 6q deletion was associated with cases that did not undergo disease progression. In fact, no case with deletion of *PRDM1* died or experienced relapse/refractory disease. *PRDM1* codes for BLIMP1 which is an important driver of B cell terminal differentiation (Shaffer *et al.*, 2002). *PRDM1* deletion has been reported previously in adult ABC-like DLBCL but not in GCB-like DLBCL (Pasqualucci *et al.*, 2006), suggesting that it would be associated with a poor prognosis. In contrast, we identified *PRDM1* in 0/18 (0.0%) cases of disease progression and 18/119 (15.1%). It is possible that this disparity in prognostic value is explained by the fact that Pascqualucci *et al.* (2006) screened primarily adult DLBCL patients while our cohort is primarily paediatric BL. The identification of a marker of good risk in paediatric B-NHL opens up the possibility for treatment de-escalation.

Our investigation of 105 endemic BL cases is the largest genomic study of this disease subtype. Recent studies have described the genetics of endemic BL without relating these findings to outcome (Grande *et al.*, 2019; Panea *et al.*, 2019). We described genomic markers detectable by copy number array that could be implicated in disease progression. Perhaps most significantly, genomic events that were significant prognostic markers of progression in the sporadic cohort were shown to be unrelated to relapse in the endemic setting, including *TP53* aberration. *TP53* abnormalities were less common in endemic BL compared to sporadic BL (47/101, 46.5% vs 44/70, 62.9%, p=0.043, Fisher exact test), but more striking was the dramatic difference in biallelic abnormalities (12/101, 11.9% vs 21/70, 30%; p=0.005, Fisher exact test) and the lack of association with disease progression. We hypothesise that this is primarily due to a different clinical setting and treatment protocols. It is likely, given this data, that the survival advantage conferred to the tumour by *TP53* aberration in sBL is important in the setting of dose intense FAB/LMB-based treatment. Treatment given to eBL patients in Malawi is

substantially less intensive (Molyneux *et al.*, 2017) and development of therapy resistance more likely, even in the absence of *TP53* aberration.

The clinical setting in Malawi provided unique difficulties to this analysis. Relatively high abscondment rates and changes in treatment protocols complicated the analysis and limited the size of the cohort that could be interrogated in survival analysis. Further sample collection is ongoing from patients diagnosed with B-NHL in Malawi, so a larger cohort is achievable in the future. Another challenge presented by the setting of the study was patient sample quality. The requirement to transport patient samples back to the UK at room temperature imposed limits on what could be achieved. Whole exome sequencing on DNA from PreservCyt-fixed samples was not possible at the point of the project's inception. Since the completion of this part of the project, other members of the lab have begun a pilot project to perform library preps on this DNA. Once this process has been fully optimised it is the plan to generate whole exome sequencing data for a large part of the Malawi cohort.

The evolution of progressive disease in paediatric B-NHL is poorly understood but is known to progress either in a late-diverging/linear pattern or an early-diverging/branching pattern. Only one study has investigated clonal evolution in BL relapses (Aukema *et al.*, 2015) and did not utilise high resolution genomic tools. In BL the majority of patients have been shown to exhibit the linear pattern of evolution. Our study of seven sBL patients with paired diagnostic and relapse material showed that cases who had biallelic *TP53* abnormalities at diagnosis followed this pattern of evolution. All six patients sBL with *TP53* mutation data had biallelic aberrations at the point of disease progression, which supports the discovery that *TP53* aberrations are a key driver of disease progression in paediatric BL. When the major diagnostic subclone harboured *TP53* biallelic aberrations the copy number profile of the relapse sample strongly resembled that of the diagnostic sample, supporting the hypothesis that clonal evolution in these cases was linear. The copy number profile was more aberrant at relapse compared to the diagnostic sample, which corroborates the findings of the karyotyping study into BL relapses (Aukema *et al.*, 2015). Increased genomic complexity is likely driven by *TP53* aberration at relapse, as we showed that cases with aberrations involving both alleles had significantly more complex genomes in Chapter 5. A significant limitation to the scope of this work is the use of log ratio as a proxy measurement for clonal fraction. This could be improved further

by the use of high-depth whole genome sequencing which would allow for analysis using bioinformatics tools such as ClonEvol to robustly identify subpopulations in paired diagnostic and relapse samples. This would address several limitations to the analysis presented in this thesis.

In summary, this PhD project has identified and characterised an important genomic prognostic marker that plays an important role in the development of treatment-resistant paediatric B-NHL. These findings may shape the way that patients are screened at diagnosis and managed through treatment. *TP53* aberrations play a major role in resistance to chemotherapy and genomic instability. Additional prognostic markers, detectable by copy number arrays, were also shown to be associated with disease progression in the paediatric setting. A good-prognostic marker, del6q, was also identified. No patients with del6q underwent disease progression. As a leading candidate within the minimal region of deletion, *PRDM1* would be amenable to clinical analysis by FISH or copy number arrays and may identify patients suitable for treatment de-escalation to reduce the toxic effects of therapy. However, to enable our findings to have significant impact in the clinic, further validation of prognostic markers is required in an additional, comparable cohort. Access to a prospective trial cohort with extensive clinical annotation would allow us to confirm the prognostic role of *TP53* aberrations in disease progression and allow us to better stratify patients based on their risk, as is done routinely in other paediatric malignancies such as B-ALL (Moorman, 2016) and medulloblastoma (Schwalbe *et al.*, 2017).

The findings of this study have given us the opportunity to ameliorate treatment-related morbidity in a large subgroup of paediatric B-NHL patients. While the introduction of rituximab to frontline therapy has all but eliminated relapsed/refractory disease, meaning that validation of *TP53* aberration as a prognostic marker cannot be validated on a prospective trial cohort. However, the control arm of the inter-B-NHL trial, who did not receive rituximab, are an ideal retrospective cohort to validate our ability to identify low-risk patients with a good-prognostic marker, such as *TP53* wildtype status, or del6q (Minard-Colin *et al.*, 2016). In addition, projects such as the ACCELERATE trial now exist to aid in the development and design of clinical trials in paediatric B cell malignancies (Pearson *et al.*, 2019). The ACCELERATE project highlights the need for combining resources worldwide to bring together all available

data and samples from patients with relapsed/refractory disease. This trial platform represents an exciting opportunity to develop prognostication in paediatric lymphoma and validate the clinical impact of *TP53* status.

The next steps to improve the outcome for patients will not simply focus on those with poor risk prognostic markers, but will heavily focus on the patients who are cured with standard therapy and have good risk markers, as the recent introduction of rituximab to frontline therapy has significantly reduced the incidence of disease progression in paediatric B-NHL (V *et al.*, 2020). Patients now experience higher-than-ever progression-free survival rates but this still comes at the cost of significant toxicity. The results presented in this thesis highlight that approximately half of the cohort without *TP53* abnormalities do not require rituximab to achieve excellent progression-free survival rates, and are candidates for treatment de-escalation. With confirmation of the findings of this project, we can significantly change how diagnosis and management of disease is performed and bring about a positive change for patient outcome.

Appendix A

Parameter	Command Line Option	Value used	Description	Default Value
Minimum and maximum cap values on analyzed data.	-cap	2	Regions with a log2 ratio greater than the cap are set to the cap value; regions with a log2 ratio less than -cap value are set to -cap. Values must be positive.	1.5
Smallest number of markers to allow in segments from the segmented data.	-js	4	Segments that contain fewer than this number of markers are joined to the neighboring segment that is closest in copy number.	4
Threshold for copy number amplifications.	-ta	0.15	Regions with a copy number gain above this value are considered amplified.	0.1
Threshold for copy number deletions.	-td	0.15	Regions with a copy number loss below this value are considered deletions.	0.1
Significance threshold for q-values.	-qvt	0.1	Regions with q-values below this number are considered significant.	0.25
Gene GISTIC	-genegistic	1	Flag indicating that the gene GISTIC algorithm should be used to calculate the significance of deletions at a gene level instead of a marker level. Allowed values= 1,0.	0
SegArray Memory Compression	-smallmem	1	Flag indicating that the SegArray memory compression scheme should be used to reduce the memory requirements of the computation for large data sets. Computation is somewhat slower with memory compression enabled. Allowed values= (1,0).	1
Broad Level Analysis	-broad	1	Flag indicating that an additional broad-level analysis should be performed. Allowed values = (1,0).	0
Two-Dimensional Quadrant Figure	-twosides	1	Flag indicating that a two-dimensional quadrant figure should be created as part of a broad analysis. Allowed values = (1,0).	0
Broad Length	-brlen	0.5	Threshold used to distinguish broad from focal events, given in units of fraction of chromosome arm.	0.98
Confidence Interval	-conf	0.99	Confidence level used to calculate the region containing a driver.	0.75
Arm Peel Off	-armpeel	1	Flag set to enable arm-level peel-off of events during peak definition. The arm-level peel-off enhancement to the arbitrated peel-off method assigns all events in the same chromosome arm of the same sample to a single	0

			peak. It is useful when peaks are split by noise or chromothripsis. Allowed values= (1,0).	
Gene Tables	-savegene	1	Flag indicating that gene tables should be saved. Allowed values= (1,0).	0
Remove Sex Chromosomes	-rx	1	Flag indicating whether to remove data from the sex chromosomes before analysis. Allowed values= (1,0).	1
Gene Collapse Method	-gcm	extreme	Method for reducing marker-level copy number data to the gene-level copy number data in the gene tables. Markers contained in the gene are used when available, otherwise the flanking marker or markers are used. Allowed values are mean, median, min, max or extreme. The extreme method chooses whichever of min or max is furthest from diploid.	mean

Appendix A - GISTIC2.0 Parameters used for Copy Number Analysis of paediatric B-NHL.

Appendix B

Variable	Time to progression (TTP)		Overall survival (OS)		
	HR (95% CI)	*p* value	HR (95% CI)	*p* value	No. cases with abnormality (%)
Univariate analysis					
7q21q22 Gain (Poirel *et al.*)	0.54 (0.16-1.9)	0.33	0.42 (0.13-1.4)	0.17	37/162 (22.8%)
13q34 Loss (Nelson *et al*., Poirel *et al*.)	1.8 (0.6-5.6)	0.29	1.4 (0.49-4.3)	0.5	19/162 (11.7%)
13q31.3 Gain (Schiffman *et al.*)	0.86 (0.28-2.6)	0.79	1.3 (0.55-3.3)	0.51	41/162 (25.3%)

Appendix B - Univariate survival analysis of three previously reported prognostic markers in a cohort of 162 paediatric B-NHL patients

Appendix C

Sample	Paired/Unpaired Sample	Location	Consequence	cDNA_position	CDS_position	Protein Change	Codons	Existing_variation	IMPACT	SYMBOL	SIFT	PolyPhen
12/854	Paired	17:78298926	missense_variant	3264/21055	3121/15624	A1041T	Gcg/Acg	rs61359568	MODERATE	RNF213	-	benign(0)
12/854	Paired	17:78305871	missense_variant	3726/21055	3583/15624	V1195M	Gtg/Atg	rs10782008	MODERATE	RNF213	-	benign(0.095)
12/854	Paired	17:78319136	missense_variant	7144/21055	7001/15624	S2334N	aGt/aAt	rs9674961,COSM4000337,COSM4000338	MODERATE	RNF213	-	benign(0)
12/854	Paired	17:78337058	missense_variant	11655/21055	11512/15624	V3838L	Gtt/Ctt	rs35332090	MODERATE	RNF213	-	benign(0.047)
12/854	Paired	17:78337584	missense_variant	11887/21055	11744/15624	E3915G	gAa/gGa	rs61740658	MODERATE	RNF213	-	benign(0.152)
14/136	Paired	17:63010694	missense_variant	1061/4922	815/1134	F272C	tTt/tGt	COSM3927582	MODERATE	GNA13	deleterious(0)	probably_damaging(0.999)
14/325	Paired	17:63010541	missense_variant	1214/4922	968/1134	Q323P	cAa/cCa	-	MODERATE	GNA13	deleterious(0)	probably_damaging(0.918)
14/325	Paired	17:78318562	missense_variant	6570/21055	6427/15624	L2143M	Ctg/Atg	-	MODERATE	RNF213	-	benign(0.218)
14/325	Paired	17:78319853	missense_variant	7861/21055	7718/15624	P2573Q	cCg/cAg	-	MODERATE	RNF213	-	probably_damaging(0.975)
14/36	Paired	17:78327370	missense_variant	10625/21055	10482/15624	S3494R	agC/agA	-	MODERATE	RNF213	-	benign(0.05)
14/400	Paired	17:63052633	stop_gained	325/4922	79/1134	Q27*	Cag/Tag	COSM982995	HIGH	GNA13	-	-
14/414	Paired	17:63049634	stop_gained	742/4922	496/1134	R166*	Cga/Tga	COSM6081726	HIGH	GNA13	-	-
14/486	Paired	17:63052633-63052642	frameshift_variant	316-325/4922	70-79/1134	EAEQ24-27X	GAGGCCGAGCag/ag	-	HIGH	GNA13	-	-
15/360	Paired	17:63052527	missense_variant	431/4922	185/1134	S62F	tCc/tTc	-	MODERATE	GNA13	deleterious(0)	probably_damaging(0.998)
15/464	Paired	17:78337083	missense_variant	11680/21055	11537/15624	R3846H	cGt/cAt	-	MODERATE	RNF213	-	benign(0.014)
16/511	Paired	17:63052529	missense_variant	429/4922	183/1134	K61N	aaG/aaC	COSM4068759	MODERATE	GNA13	deleterious(0)	probably_damaging(0.999)
16/611	Paired	17:78320062	missense_variant	8070/21055	7927/15624	R2643S	Cgc/Agc	COSM5454625,COSM5454626	MODERATE	RNF213	-	probably_damaging(0.999)
21/153	Paired	17:63052633	stop_gained	325/4922	79/1134	Q27*	Cag/Tag	COSM982995	HIGH	GNA13	-	-
21/228	Paired	17:78357486	missense_variant	14223/21055	14080/15624	K4694Q	Aaa/Caa	-	MODERATE	RNF213	-	possibly_damaging(0.796)
21/322	Paired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	-	benign(0.001)

5/796	Paired	17:63010556	missense_variant	1199/4922	953/1134	C318F	tGc/tTc	-	MODERATE	GNA13	deleterious(0.02)	benign(0.276)
10/10065	Unpaired	17:78263486	missense_variant	1105/21055	962/1562 4	M321T	aTg/aCg	rs17853989	MODERATE	RNF213	tolerated(0.73)	benign(0)
10/10065	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
10/134	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
10/21	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
10/28	Unpaired	17:78263486	missense_variant	1105/21055	962/1562 4	M321T	aTg/aCg	rs17853989	MODERATE	RNF213	tolerated(0.73)	benign(0)
10/45	Unpaired	17:63052630	stop_gained	328/4922	82/1134	Q28*	Caa/Taa	-	HIGH	GNA13	-	-
10/45	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
11/581	Unpaired	17:78263486	missense_variant	1105/21055	962/1562 4	M321T	aTg/aCg	rs17853989	MODERATE	RNF213	tolerated(0.73)	benign(0)
11/581	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
11/602	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
11/656	Unpaired	17:78263486	missense_variant	1105/21055	962/1562 4	M321T	aTg/aCg	rs17853989	MODERATE	RNF213	tolerated(0.73)	benign(0)
11/656	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
18/721	Unpaired	17:63010916	missense_variant	839/4922	593/1134	L198P	cTt/cCt	-	MODERATE	GNA13	tolerated(0.16)	probably_damaging(0.983)
18/721	Unpaired	17:78263486	missense_variant	1105/21055	962/1562 4	M321T	aTg/aCg	rs17853989	MODERATE	RNF213	tolerated(0.73)	benign(0)
18/721	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
12/195	Unpaired	17:78263486	missense_variant	1105/21055	962/1562 4	M321T	aTg/aCg	rs17853989	MODERATE	RNF213	tolerated(0.73)	benign(0)
12/195	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
12/502	Unpaired	17:78263486	missense_variant	1105/21055	962/1562 4	M321T	aTg/aCg	rs17853989	MODERATE	RNF213	tolerated(0.73)	benign(0)
12/502	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
12/760	Unpaired	17:63014381	missense_variant	797/4922	551/1134	L184P	cTt/cCt	COSM1159762	MODERATE	GNA13	deleterious(0)	probably_damaging(0.96)
12/760	Unpaired	17:63014381	missense_variant	797/4922	551/1134	L184R	cTt/cGt	COSM1159762	MODERATE	GNA13	deleterious(0)	probably_damaging(0.914)
12/760	Unpaired	17:78263486	missense_variant	1105/21055	962/1562 4	M321T	aTg/aCg	rs17853989	MODERATE	RNF213	tolerated(0.73)	benign(0)
12/760	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)

14/165	Unpaired	17:78263486	missense_variant	1105/21055	962/1562 4	M321T	aTg/aCg	rs17853989	MODERATE	RNF213	tolerated(0.73)	benign(0)
14/165	Unpaired	17:78333985	stop_gained	11322/21055	11179/15624	Q3727*	Cag/Tag	-	HIGH	RNF213	-	-
14/167	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
14/299	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
14/334	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
14/334	Unpaired	17:78350165	missense_variant	13393/21055	13250/15624	R4417H	cGt/cAt	rs150148627	MODERATE	RNF213	tolerated(0.33)	benign(0)
15/156	Unpaired	17:63010848	missense_variant	907/4922	661/1134	V221L	Gtt/Ctt	rs1062597	MODERATE	GNA13	deleterious(0)	benign(0.065)
15/156	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
15/338	Unpaired	17:78263486	missense_variant	1105/21055	962/1562 4	M321T	aTg/aCg	rs17853989	MODERATE	RNF213	tolerated(0.73)	benign(0)
15/338	Unpaired	17:78272230	missense_variant	2265/21055	2122/15624	H708D	Cac/Gac	rs72849837	MODERATE	RNF213	tolerated(0.61)	benign(0)
15/338	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
15/353	Unpaired	17:78263486	missense_variant	1105/21055	962/1562 4	M321T	aTg/aCg	rs17853989	MODERATE	RNF213	tolerated(0.73)	benign(0)
16/772	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
16/772	Unpaired	17:78313251	missense_variant	5227/21055	5084/15624	R1695Q	cGa/cAa	rs370463812	MODERATE	RNF213	deleterious(0)	probably_damaging(0.995)
16/816	Unpaired	17:78263486	missense_variant	1105/21055	962/1562 4	M321T	aTg/aCg	rs17853989	MODERATE	RNF213	tolerated(0.73)	benign(0)
16/816	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
16/817	Unpaired	17:63014390	stop_gained	788/4922	542/1134	L181*	tTg/tAg	COSM4170816	HIGH	GNA13	-	-
16/817	Unpaired	17:78263486	missense_variant	1105/21055	962/1562 4	M321T	aTg/aCg	rs17853989	MODERATE	RNF213	tolerated(0.73)	benign(0)
16/817	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
16/827	Unpaired	17:78263486	missense_variant	1105/21055	962/1562 4	M321T	aTg/aCg	rs17853989	MODERATE	RNF213	tolerated(0.73)	benign(0)
16/827	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
16/827	Unpaired	17:78355462	missense_variant	14056/21055	13913/15624	T4638I	aCc/aTc	rs141301945	MODERATE	RNF213	deleterious(0.01)	possibly_damaging(0.873)
20/3020	Unpaired	17:78263486	missense_variant	1105/21055	962/1562 4	M321T	aTg/aCg	rs17853989	MODERATE	RNF213	tolerated(0.73)	benign(0)
20/3020	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)

20/4012	Unpaired	17:63010775	missense_variant	980/4922	734/1134	F245S	tTc/tCc	COSM220991	MODERATE	GNA13	deleterious(0)	benign(0.405)
20/4012	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
21/199	Unpaired	17:63052633	stop_gained	325/4922	79/1134	Q27*	Cag/Tag	COSM982995	HIGH	GNA13	-	-
21/199	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
21/257	Unpaired	17:78247124	missense_variant	325/21055	182/1562 4	P61L	cCg/cTg	rs9913317	MODERATE	RNF213	tolerated(0.53)	benign(0.007)
21/257	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
21/37	Unpaired	17:78263486	missense_variant	1105/21055	962/1562 4	M321T	aTg/aCg	rs17853989	MODERATE	RNF213	tolerated(0.73)	benign(0)
21/55	Unpaired	17:78263486	missense_variant	1105/21055	962/1562 4	M321T	aTg/aCg	rs17853989	MODERATE	RNF213	tolerated(0.73)	benign(0)
21/55	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
21/55	Unpaired	17:78313281	missense_variant	5257/21055	5114/15624	T1705K	aCg/aAg	rs147868237,COSM148363	MODERATE	RNF213	deleterious(0)	probably_damaging(0.982)
22/22	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
22/9	Unpaired	17:78263486	missense_variant	1105/21055	962/1562 4	M321T	aTg/aCg	rs17853989	MODERATE	RNF213	tolerated(0.73)	benign(0)
22/9	Unpaired	17:78298926	missense_variant	3264/21055	3121/15624	A1041T	Gcg/Acg	rs61359568	MODERATE	RNF213	tolerated(0.58)	benign(0)
22/9	Unpaired	17:78306280	missense_variant	4135/21055	3992/15624	D1331G	gAc/gGc	rs8074015	MODERATE	RNF213	tolerated(0.17)	benign(0.001)
22/9	Unpaired	17:78357600	missense_variant	14337/21055	14194/15624	K4732E	Aaa/Gaa	rs12944385	MODERATE	RNF213	tolerated(0.65)	benign(0.033)
16/818	Unpaired	17:78263486	missense_variant	1105/21055	962/1562 4	M321T	aTg/aCg	rs17853989	MODERATE	RNF213	tolerated(0.73)	benign(0)

Appendix C - Detailed annotation for mutations identified on 17q involving GNA13 and RNF213

Appendix D

Sample	Paired/Unpaired Sample	Location	Consequence	cDNA_position	CDS_position	Protein Change	Codons	Existing_variation	IMPACT	SYMBOL	SIFT	PolyPhen
12/816	Paired	3:193207589	missense_variant	991/4208	668/3591	K223T	aAg/aCg	-	MODERATE	ATP13A4	tolerated(0.21)	benign(0.054)
12/816	Paired	3:195516343	missense_variant	2568/17110	2108/16239	G703V	gGt/gTt	COSM3992972	MODERATE	MUC4	deleterious_low_confidence(0)	probably_damaging(0.955)
12/854	Paired	3:193209178	missense_variant	866/4208	543/3591	I181M	atA/atG	rs6788448	MODERATE	ATP13A4	deleterious(0.04)	probably_damaging(0.989)
12/854	Paired	3:193361167	missense_variant	1491/6439	1257/2994	I419M	atA/atG	rs143319805,CM080464	MODERATE	OPA1	deleterious(0)	probably_damaging(0.997)
14/36	Paired	3:193355812	missense_variant	1287/6439	1053/2994	F351L	ttC/ttA	-	MODERATE	OPA1	deleterious(0)	probably_damaging(0.993)
14/477	Paired	3:193332589	stop_gained	344/6439	110/2994	S37*	tCa/tAa	COSM729922	HIGH	OPA1	-	-
14/486	Paired	3:193355814	missense_variant	1289/6439	1055/2994	P352Q	cCa/cAa	-	MODERATE	OPA1	deleterious(0)	probably_damaging(1)
16/488	Paired	3:195518132	missense_variant	779/17110	319/16239	V107L	Gta/Tta	-	MODERATE	MUC4	tolerated_low_confidence(0.34)	benign(0.101)
21/224	Paired	3:196530036	splice_donor_variant	-	-		-	rs753970309,COSM5494668	HIGH	PAK2	-	-
21/224	Paired	3:196530037	splice_donor_variant	-	-		-	rs757545981,COSM3660448	HIGH	PAK2	-	-
21/292	Paired	3:193355802	missense_variant	1277/6439	1043/2994	A348D	gCt/gAt	rs398124302,CD012270	MODERATE	OPA1	deleterious(0)	probably_damaging(1)
21/322	Paired	3:196529902	missense_variant	625/6139	303/1575	Q101H	caG/caC	rs201465227	MODERATE	PAK2	deleterious(0.02)	possibly_damaging(0.886)
5/894	Paired	3:196529982	missense_variant	705/6139	383/1575	K128R	aAg/aGg	rs78043821	MODERATE	PAK2	tolerated(0.17)	benign(0.41)
10/134	Unpaired	3:195453014	missense_variant	1666/2589	1540/2130	G514R	Ggg/Agg	rs3828410,COSM4002561	MODERATE	MUC20	tolerated_low_confidence(0.7)	benign(0)
10/134	Unpaired	3:195456545	missense_variant	2122/2589	1996/2130	R666W	Cgg/Tgg	rs11923495,COSM1421838,COSM1421839	MODERATE	MUC20	deleterious(0.02)	probably_damaging(0.999)
10/21	Unpaired	3:193036783	missense_variant	2148/4027	2030/3657	A677V	gCc/gTc	rs73888252,COSM6861920	MODERATE	ATP13A5	tolerated(0.1)	benign(0.015)
10/28	Unpaired	3:193036783	missense_variant	2148/4027	2030/3657	A677V	gCc/gTc	rs73888252,COSM6861920	MODERATE	ATP13A5	tolerated(0.1)	benign(0.015)
10/45	Unpaired	3:194147859	missense_variant	3862/7720	3070/3681	G1024S	Ggt/Agt	rs199767870,COSM138415	MODERATE	ATP13A3	deleterious(0.02)	benign(0.04)
11/581	Unpaired	3:193036783	missense_variant	2148/4027	2030/3657	A677V	gCc/gTc	rs73888252,COSM6861920	MODERATE	ATP13A5	tolerated(0.1)	benign(0.015)
12/195	Unpaired	3:195453014	missense_variant	1666/2589	1540/2130	G514R	Ggg/Agg	rs3828410,COSM4002561	MODERATE	MUC20	tolerated_low_confidence(0.7)	benign(0)
12/502	Unpaired	3:194790820	missense_variant	915/2727	806/1182	R269L	cGc/cTc	COSM6223684,COSM6223685,COSM6223686	MODERATE	XXYLT1	deleterious(0)	probably_damaging(1)

12/502	Unpaired	3:195453014	missense_variant	1666/2589	1540/2130	G514R	Ggg/Agg	rs3828410,COSM4002561	MODERATE	MUC20	tolerated_low_confidence(0.7)	benign(0)
12/760	Unpaired	3:193036783	missense_variant	2148/4027	2030/3657	A677V	gCc/gTc	rs73888252,COSM6861920	MODERATE	ATP13A5	tolerated(0.1)	benign(0.015)
12/760	Unpaired	3:195453014	missense_variant	1666/2589	1540/2130	G514R	Ggg/Agg	rs3828410,COSM4002561	MODERATE	MUC20	tolerated_low_confidence(0.7)	benign(0)
12/77	Unpaired	3:195453014	missense_variant	1666/2589	1540/2130	G514R	Ggg/Agg	rs3828410,COSM4002561	MODERATE	MUC20	tolerated_low_confidence(0.7)	benign(0)
12/77	Unpaired	3:196842949	missense_variant	1581/5034	1391/2781	P464R	cCt/cGt	rs772348654,COSM70455	MODERATE	DLG1	deleterious(0)	probably_damaging(0.996)
14/167	Unpaired	3:193036783	missense_variant	2148/4027	2030/3657	A677V	gCc/gTc	rs73888252,COSM6861920	MODERATE	ATP13A5	tolerated(0.1)	benign(0.015)
14/334	Unpaired	3:193036783	missense_variant	2148/4027	2030/3657	A677V	gCc/gTc	rs73888252,COSM6861920	MODERATE	ATP13A5	tolerated(0.1)	benign(0.015)
16/772	Unpaired	3:192635514	missense_variant	437/3368	116/1476	L39R	cTc/cGc	-	MODERATE	MB21D2	deleterious(0)	benign(0.259)
16/817	Unpaired	3:193036783	missense_variant	2148/4027	2030/3657	A677V	gCc/gTc	rs73888252,COSM6861920	MODERATE	ATP13A5	tolerated(0.1)	benign(0.015)
16/817	Unpaired	3:195456545	missense_variant	2122/2589	1996/2130	R666W	Cgg/Tgg	rs11923495,COSM1421838,COSM1421839	MODERATE	MUC20	deleterious(0.02)	probably_damaging(0.999)
16/827	Unpaired	3:193036783	missense_variant	2148/4027	2030/3657	A677V	gCc/gTc	rs73888252,COSM6861920	MODERATE	ATP13A5	tolerated(0.1)	benign(0.015)
21/199	Unpaired	3:193036783	missense_variant	2148/4027	2030/3657	A677V	gCc/gTc	rs73888252,COSM6861920	MODERATE	ATP13A5	tolerated(0.1)	benign(0.015)
21/199	Unpaired	3:195453014	missense_variant	1666/2589	1540/2130	G514R	Ggg/Agg	rs3828410,COSM4002561	MODERATE	MUC20	tolerated_low_confidence(0.7)	benign(0)
21/257	Unpaired	3:193036783	missense_variant	2148/4027	2030/3657	A677V	gCc/gTc	rs73888252,COSM6861920	MODERATE	ATP13A5	tolerated(0.1)	benign(0.015)
21/37	Unpaired	3:193036783	missense_variant	2148/4027	2030/3657	A677V	gCc/gTc	rs73888252,COSM6861920	MODERATE	ATP13A5	tolerated(0.1)	benign(0.015)
21/37	Unpaired	3:195486072	missense_variant	15369/17110	14909/16239	T4970M	aCg/aTg	rs114608303,COSM3357692,COSM3357693,COSM3357694	MODERATE	MUC4	tolerated_low_confidence(0.08)	benign(0.007)
22/22	Unpaired	3:193036783	missense_variant	2148/4027	2030/3657	A677V	gCc/gTc	rs73888252,COSM6861920	MODERATE	ATP13A5	tolerated(0.1)	benign(0.015)
22/22	Unpaired	3:195453014	missense_variant	1666/2589	1540/2130	G514R	Ggg/Agg	rs3828410,COSM4002561	MODERATE	MUC20	tolerated_low_confidence(0.7)	benign(0)

Appendix D - Detailed annotation for mutations idenfied on 3q29

Appendix E

ID	Diagnosis	CNA	Size (Mb)	% of 17p	Mutation	Nucleotide	VAF (%)	Activity*	Functional classification**	Mutation type	TP53 status
1/49	BL	no abnormality			no mutation					no mutation	no abnormality
10/10065 F1	BL	17p Loss	22.2	99.6%	G266E	G>A	85	Inactive	non-functional	Homozygous	Biallelic
10/134	DLBCL	no abnormality			no mutation					no mutation	no abnormality
10/21	BL	no abnormality			R249T	G>C	21	Inactive	non-functional	Heterozygous	Monoallelic
10/28	BL	17p Loss	20.5	91.9%	no mutation					no mutation	Monoallelic
10/45	BL	no abnormality			no mutation					no mutation	no abnormality
11/581	BL	no abnormality			D208G	A>G	39	Fully active	functional	Heterozygous	Monoallelic
11/602	BNHL	no abnormality			E286K	G>A	41	Inactive	non-functional	Heterozygous	Monoallelic
11/634	BL	no abnormality			no mutation					no mutation	no abnormality
11/656	BL	no abnormality			S241P	T>C	38	Inactive	non-functional	Heterozygous	Monoallelic
12/195	BNHL	no abnormality			no mutation					no mutation	no abnormality
12/502	DLBCL	no abnormality			no mutation					no mutation	no abnormality
12/510	BL	no abnormality			no mutation					no mutation	no abnormality
12/709 F1	BL	no abnormality			no mutation					no mutation	no abnormality
12/760	BL	no abnormality			no mutation					no mutation	no abnormality
12/77	BLL-11q	no abnormality			no mutation					no mutation	no abnormality
12/816	BLL-11q	no abnormality			no mutation					no mutation	no abnormality
12/837	BNHL	no abnormality			no mutation					no mutation	no abnormality
12/854	BLL-11q	no abnormality			no mutation					no mutation	no abnormality
12/91	BL	no abnormality			L289fs	TC>*	89	The activity of truncated p53 is assumed to be nil	frameshift	Homozygous	Biallelic
14/136	BNHL	no abnormality			C176F	G>T	40	Inactive	partially functional	Heterozygous	Monoallelic
14/165	BL	no abnormality			R248Q	G>A	44	Inactive	non-functional	Heterozygous	Monoallelic
14/167	BL	no abnormality			no mutation					no mutation	no abnormality

14/228	BL	no abnormality			no mutation					no mutation	no abnormality
14/299	BL	no abnormality			E258D	A>C	48	Inactive	non-functional	Heterozygous	Monoallelic
14/325	BL	no abnormality			R158H, R175H	G>A, G>A	54, 36	Inactive, Inactive	non-functional, non-functional	Compound Heterozygous	Biallelic
14/334	DLBCL	17p Loss	22.2	99.6%	no mutation					no mutation	Monoallelic
14/36	BL	no abnormality			no mutation					no mutation	no abnormality
14/400	BL	17p Loss	19.7	88.3%	V122fs	Gdel	40	The activity of truncated p53 is assumed to be nil	frameshift	Hemizygous	Biallelic
14/414	BL	no abnormality			R175H, G245S	G>A, G>A	49, 36	Inactive, Inactive	non-functional, non-functional	Compound Heterozygous	Biallelic
14/477	DLBCL	no abnormality			R248W	C>T	49	Inactive	non-functional	Heterozygous	Monoallelic
14/486	BL	no abnormality			V173M	G>A	36	Inactive	non-functional	Heterozygous	Monoallelic
14/502	DLBCL	no abnormality			no mutation					no mutation	no abnormality
15/156	BNHL	no abnormality			S241Y, G245S	C>A, G>A	32, 22	Inactive, Inactive	non-functional, non-functional	Compound Heterozygous	Biallelic
15/268	BL	no abnormality			no mutation					no mutation	no abnormality
15/310	BL	no abnormality			G245S	G>A	39	Inactive	non-functional	Heterozygous	Monoallelic
15/314	BL	no abnormality			R158H, P278S	G>A, C>T	53, 39	Inactive, Inactive	non-functional, non-functional	Compound Heterozygous	Biallelic
15/338	BL	no abnormality			no mutation					no mutation	no abnormality
15/353 F1	BL	no abnormality			R248Q, R273C	G>A, C>T	44, 20	Inactive, Inactive	non-functional, non-functional	Compound Heterozygous	Biallelic
15/360	DLBCL	no abnormality			no mutation					no mutation	no abnormality
15/461	BL	no abnormality			no mutation					no mutation	no abnormality
15/464	BL	no abnormality			Y205F	A>T	39	The activity of truncated p53 is assumed to be nil	non-functional	Heterozygous	Monoallelic
15/476	BL	no abnormality			no mutation					no mutation	no abnormality
15/526	BL	no abnormality			no mutation					no mutation	no abnormality
16/391 F1	BL	17p CNN-LOH	22.2	99.6%	R248Q	G>A	40	Inactive	non-functional	Homozygous	Biallelic
16/488	BL	no abnormality			G266E	G>A	39	Inactive	non-functional	Heterozygous	Monoallelic
16/511	BL	no abnormality			no mutation					no mutation	no abnormality

16/547 F1	DLBCL	17p CNN-LOH	22.3	100.0 %	no mutation					no mutation	Monoallelic
16/611	BL	no abnormality			no mutation					no mutation	no abnormality
16/772	DLBCL	no abnormality			no mutation					no mutation	no abnormality
16/816	BL	17p Loss	20.6	92.4%	R175H	G>A	31	Inactive	non-functional	Hemizygous	Biallelic
16/817	BL	no abnormality			no mutation					no mutation	no abnormality
16/818 F1	BL	17p Loss	17.9	80.3%	V216G	T>G	92	Inactive	non-functional	Hemizygous	Biallelic
16/827	BL	17p Loss	20.6	92.4%	D281G	A>G	24	Inactive	non-functional	Hemizygous	Biallelic
16/847	DLBCL	17p Loss	21.5	96.4%	R248Q	G>A	56	Inactive	non-functional	Hemizygous	Biallelic
18/721 P1	BL	no abnormality			no mutation					no mutation	no abnormality
20/3020	DLBCL	17p Loss	19	85.2%	no mutation					no mutation	Monoallelic
20/4012	BL	no abnormality			no mutation					no mutation	no abnormality
21/153	BNHL	no abnormality			Y126S	A>C	96	Inactive	non-functional	Homozygous	Biallelic
21/199	BL	no abnormality			R282W	C>T	18	Inactive	non-functional	Heterozygous	Monoallelic
21/224	DLBCL	no abnormality			no mutation					no mutation	no abnormality
21/228	DLBCL	no abnormality			no mutation					no mutation	no abnormality
21/257	BL	no abnormality			Y126H, G245S	T>C, G>A	47, 42	Inactive, Inactive	non-functional, non-functional	Compound Heterozygous	Biallelic
21/292	DLBCL	no abnormality			no mutation					no mutation	no abnormality
21/322	BL	17p Loss	22.2	99.6%	R273H	G>A	59	Inactive	non-functional	Hemizygous	Biallelic
21/37	BL	no abnormality			R248Q	G>A	41	Inactive	non-functional	Heterozygous	Monoallelic
21/55	BL	no abnormality			C135R	T>C	38	Inactive	non-functional	Heterozygous	Monoallelic
22/22	BL	no abnormality			R175H	G>A	50	Inactive	non-functional	Heterozygous	Monoallelic
22/9	BL	17p CNN-LOH	14	62.8%	T155N	C>A	84	Inactive	non-functional	Homozygous	Biallelic
3/95	DLBCL	no abnormality			no mutation					no mutation	no abnormality
4/145	BLL-11q	no abnormality			no mutation					no mutation	no abnormality
4/17	BNHL	no abnormality			no mutation					no mutation	no abnormality
4/271	DLBCL	no abnormality			no mutation					no mutation	no abnormality

4/281	BL	17p CNN-LOH	21.8	97.8%	G245S	G>A	88, 90	Inactive	non-functional	Homozygous	Biallelic
4/283	DLBCL	17p Loss	22.2	99.6%	R213Q	G>A	71	Inactive	non-functional	Hemizygous	Biallelic
4/297	BL	no abnormality			no mutation					no mutation	no abnormality
4/349	BL	no abnormality			no mutation					no mutation	no abnormality
4/438	DLBCL	no abnormality			no mutation					no mutation	no abnormality
4/830	BNHL	17p Loss	22.2	99.6%	L257P, R282W	T>C, C>T	65, 30	Inactive, Inactive	non-functional	Hemizygous	Biallelic
4/911	BL	17p CNN-LOH	17.2	77.1%	R175H	G>A	91	Inactive	non-functional	Homozygous	Biallelic
5/212	DLBCL	no abnormality			no mutation					no mutation	no abnormality
5/283	BL	no abnormality			R333P	G>C	36	Partial activity	partially functional	Heterozygous	Monoallelic
5/311	BL	no abnormality			no mutation					no mutation	no abnormality
5/437	BL	17p Loss	22.2	99.6%	R196P	G>C	88	Inactive	non-functional	Hemizygous	Biallelic
5/685	BL	17p CNN-LOH	12.9	57.8%	no mutation					no mutation	Monoallelic
5/686	BL	no abnormality			T256K, R248Q	C>A, G>A	40, 41	Inactive, Inactive	partially functional, non-functional	Compound Heterozygous	Biallelic
5/790	BL	no abnormality			R273C	C>T	45	Inactive	non-functional	Heterozygous	Monoallelic
5/796	BL	no abnormality			no mutation					no mutation	no abnormality
5/894 P1	DLBCL	no abnormality			G187S	G>A	39	Fully active	functional	Heterozygous	Monoallelic
5/915	BL	17p Loss	18.5	83.0%	no mutation					no mutation	Monoallelic
5/96	BL	no abnormality			no mutation					no mutation	no abnormality
6/475	BL	17p Loss	19.4	87.0%	R273C	C>T	86	Inactive	non-functional	Hemizygous	Biallelic
7/146	BL	no abnormality			P151A	C>G	20	Inactive	non-functional	Heterozygous	Monoallelic
7/550	BL	17p CNN-LOH	22.2	99.6%	R175H	G>A	92	Inactive	non-functional	Homozygous	Biallelic
9/25	BL	17p CNN-LOH	16.4	73.5%	H179R	A>G	92	Inactive	non-functional	Homozygous	Biallelic

Appendix E - Overview of TP53 abnormalities identified in paediatric B-NHL patient samples at the time of initial diagnosis

Appendix F

	Time to progression (TTP)		Overall survival (OS)		
	HR (95% CI)	p value	HR (95% CI)	p value	No. cases with abnormality (%)
CNS involvement	1.53 (0.2-12.3)	0.686	4.40 (0.9-22.0)	0.072	4/47 (8.5%)*
BM involvement	1.68 (0.5-6.0)	0.421	3.40 (0.9-13.0)	0.067	14/48 (29.2%)*
TP53 deletion	1.60 (0.4-6.0)	0.470	0.44 (0.1-3.4)	0.430	10 (17.2%)
TP53 CNN-LOH	4.80 (1.4-16.0)	0.011	3.40 (0.9-13.0)	0.069	6 (10.3%)
TP53 mutation	8.30 (1.1-64.0)	0.043	7.20 (0.9-56.0)	0.060	35 (60.3%)
Any TP53 abnormality	7.00 (0.9-54.0)	0.062	6.20 (0.8-48.0)	0.083	37 (63.8%)
TP53 monoallelic abnormality	4.40 (0.5-42.0)	0.200	7.10 (0.8-61.0)	0.074	16 (27.6%)
TP53 biallelic abnormality	8.80 (1.1-70.0)	0.041	5.20 (0.6-45.0)	0.130	21 (36.2%)

*Appendix F - Survival analysis of TP53 abnormalities in subsets of the paediatric B-NHL cohort. * Value calculated as a percentage of cases with available data.*

Appendix G

B number	Treatment (at diagnosis)	Sex	Age at Diagnosis	Disease Stage	CSF involvement	BM involvement	MYC Translocation	EBER Status	Included in RR Analysis	Included in OS Analysis	Sample Batch	17p CNN-LOH	17q CNN-LOH	17p Loss	3q29 Gain	MCL1 Gain
B095	NA	M	10	NA	NA	NA	MYC/IGL	1	No	No	1	No	Yes	No	No	No
B105	L (BL Protocol)	M	3	1	0	0	IGH/MYC	1	No	Yes	1	No	No	No	No	No
B121	L (BL Protocol)	M	6	2	0	0	MYC/IGL	1	Yes	Yes	1	No	No	No	No	No
B150	NL (BL Protocol 1)	M	12	3	0	0	IGH/MYC	1	Yes	Yes	1	No	No	No	No	No
B153	NL (BL Protocol 1)	F	5	4	1	0	IGH/MYC	1	Yes	Yes	1	No	Yes	No	Yes	Yes
B156	NL (BL Protocol 1)	F	9	2	0	0	IGH/MYC	NA	No	No	1	No	Yes	No	Yes	No
B160	NL (BL Protocol 1)	M	12	4	1	1	IGH/MYC	1	Yes	Yes	1	No	No	Yes	No	Yes
B172	NL (BL Protocol 1)	F	3	4	1	1	IGH/MYC	1	Yes	Yes	1	No	No	Yes	No	Yes
B174	NL (BL Protocol 1)	M	6	1	0	0	IGH/MYC	NA	Yes	Yes	1	No	No	Yes	No	Yes
B186	NL (Protocol 2) no dox	M	5	2	0	0	MYC/IGL	1	Yes	Yes	1	No	No	Yes	No	No
B190	NL (Protocol 2) no dox	F	11	1	0	0	IGH/MYC	1	Yes	Yes	1	No	No	No	No	No
B195	NL (Protocol 3) + dox	F	5	3	NA	NA	IGH/MYC	1	Yes	Yes	1	No	No	No	No	No
B201	NL (Protocol 3) + dox	M	4	4	NA	0	IGH/MYC	1	Yes	Yes	1	No	No	No	No	No
B210	NA	M	7	NA	NA	NA	MYC/IGL	1	No	No	1	No	No	No	No	Yes
B216	NA	M	2	NA	NA	NA	IGH/MYC	1	No	No	1	No	Yes	No	No	Yes
B232	L (BL Protocol)	M	12	2	0	0	IGH/MYC	1	Yes	Yes	1	No	Yes	Yes	No	Yes
B264	NL (Protocol 3) + dox	M	8	3	0	NA	IGH/MYC	NA	Yes	Yes	1	No	No	No	No	No
B272	NA	F	6	1	0	0	IGH/MYC	1	No	No	1	No	No	No	Yes	No
B273	NL (Protocol 3) + dox	F	13	4	1	0	MYC/IGL	1	Yes	Yes	1	No	No	No	No	No
B282	NL (Protocol 3) + dox	M	5	4	1	0	IGH/MYC	1	Yes	Yes	1	Yes	No	No	No	No
B284	NA	F	5	4	NA	NA	MYC/IGK	1	No	No	1	No	No	No	No	No

B285	NL (Protocol 3) + dox	F	4	3	0	0	IGH/MYC	1	Yes	Yes	1	No	No	Yes	No	No
B286	NL (Protocol 2) no dox	M	6	2	NA	NA	IGH/MYC	1	Yes	Yes	1	No	Yes	No	No	Yes
B287	NL (Protocol 3) + dox	M	5	4	1	NA	IGH/MYC	1	Yes	Yes	1	No	No	No	No	No
B288	NA	F	9	3	NA	NA	IGH/MYC	1	No	No	1	No	No	No	No	No
B289	NL (Protocol 3) + dox	F	7	3	NA	NA	IGH/MYC	1	Yes	Yes	1	No	Yes	No	No	No
B294	NL (Protocol 3) + dox	F	6	3	0	NA	IGH/MYC	1	Yes	Yes	1	No	No	No	No	No
B298	NL (Protocol 3) + dox	M	6	3	0	0	MYC/IGK	1	Yes	Yes	1	No	Yes	No	No	No
B309	NL (Protocol 3) + dox	F	6	4	NA	1	IGH/MYC	NA	No	No	1	No	Yes	No	No	No
B311	NL (Protocol 3) + dox	F	8	2	0	0	IGH/MYC	1	Yes	Yes	1	No	No	No	No	No
B313	NA	F	4	3	0	NA	MYC/IGL	1	No	No	1	No	No	No	No	No
B314	NL (Protocol 3) + dox	F	4	4	0	0	IGH/MYC	1	Yes	Yes	1	No	No	No	No	No
B321	NL (Protocol 3) + dox	M	3	3	0	NA	MYC/IGL	1	Yes	Yes	1	No	No	No	No	No
B324	NA	F	3	NA	NA	NA	IGH/MYC	1	No	No	1	No	No	No	No	Yes
B326	NL (Protocol 3) + dox	F	6	3	NA	NA	IGH/MYC	1	Yes	Yes	1	No	No	No	No	No
B330	NL (Protocol 2) no dox	M	6	2	0	0	IGH/MYC	1	Yes	Yes	1	No	No	No	No	No
B343	NL (Protocol 3) + dox	M	10	1	0	NA	IGH/MYC	NA	Yes	Yes	1	No	No	No	No	No
B346	NL (Protocol 3) + dox	M	7	3	0	NA	IGH/MYC	1	Yes	Yes	1	No	No	No	Yes	No
B347	NL (Protocol 3) + dox	M	7	3	NA	NA	IGH/MYC	NA	Yes	Yes	1	No	No	No	Yes	No
B349	NL (Protocol 3) + dox	F	6	3	0	NA	IGH/MYC	1	Yes	Yes	1	No	No	No	No	Yes
B361	NL (Protocol 3) + dox	M	9	3	0	0	IGH/MYC	NA	No	No	1	No	No	No	No	No
B363	L (BL Protocol)	F	12	1	0	0	MYC/IGL	1	Yes	Yes	1	No	No	Yes	No	No
B370	NL (Protocol 3) + dox	M	8	3	NA	NA	IGH/MYC	1	Yes	Yes	1	No	No	No	No	No
B371	NL (Protocol 3) + dox	M	10	3	0	0	IGH/MYC	1	Yes	Yes	1	No	No	No	No	Yes
B374	NL (Protocol 2) no dox	M	5	1	0	0	IGH/MYC	1	Yes	Yes	1	No	No	No	No	Yes
B375	NL (Protocol 3) + dox	M	8	3	0	0	MYC/IGL	1	Yes	Yes	1	No	No	No	No	No

B389	NL (Protocol 3) + dox	F	11	NA	NA	NA	IGH/MYC	1	No	No	1	No	No	No	No	No
B391	NL (Protocol 3) + dox	M	8	4	0	1	MYC/IGK	1	Yes	Yes	1	No	No	No	No	No
B397	NL (Protocol 3) + dox	F	10	3	0	0	MYC/IGL	1	Yes	Yes	1	No	No	No	No	No
B409	NL (Protocol 3) + dox	M	3	NA	NA	NA	IGH/MYC	NA	No	No	1	No	No	No	No	No
B410	NL (Protocol 3) + dox	M	3	3	NA	NA	IGH/MYC	1	No	No	1	No	No	No	No	No
B413	NL (Protocol 3) + dox	M	12	4	NA	1	IGH/MYC	1	Yes	Yes	1	No	No	No	No	No
B420	NA	M	6	NA	NA	NA	MYC/IGL	NA	No	No	1	No	No	No	No	Yes
B422	NL (Protocol 3) + dox	M	5	3	0	NA	IGH/MYC	1	Yes	Yes	1	No	No	Yes	Yes	No
B423	NL (Protocol 3) + dox	M	6	3	NA	NA	IGH/MYC	1	Yes	Yes	1	No	No	No	No	No
B428	NL (Protocol 2) no dox	M	7	2	0	0	IGH/MYC	1	Yes	Yes	1	No	No	No	No	No
B436	NL (Protocol 3) + dox	M	7	3	0	0	IGH/MYC	1	Yes	Yes	1	No	Yes	No	No	No
B453	NA	NA	NA	NA	1	NA	IGH/MYC	1	No	No	1	Yes	Yes	No	No	No
B459	NA	F	9	3	NA	NA	IGH/MYC	1	No	No	1	No	No	No	No	Yes
B463	NL (Protocol 3) + dox	M	8	3	0	0	IGH/MYC	1	Yes	Yes	1	No	No	No	No	No
B483	NL (Protocol 3) + dox	M	9	4	0	0	IGH/MYC	NA	Yes	Yes	2	No	No	No	Yes	No
B492	NL (Protocol 3) + dox	M	11	1	0	0	IGH/MYC	NA	Yes	Yes	2	No	Yes	No	No	Yes
B495	NL (Protocol 3) + dox	F	7	3 or 4	0	0	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	No
B498	NL (Protocol 3) + dox	F	5	1	0	0	IGH/MYC	NA	Yes	Yes	2	No	Yes	No	No	No
B501	NL (Protocol 3) + dox	F	10	3	0	0	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	No
B505	NL (Protocol 3) + dox	M	5	1	0	0	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	No
B509	NL (Protocol 3) + dox	M	10	3	0	0	IGH/MYC	NA	Yes	Yes	2	No	Yes	No	No	No
B516	NL (Protocol 3) + dox	M	9	3	0	0	MYC/IGL	NA	Yes	Yes	2	No	No	No	No	Yes
B520	NL (Protocol 3) + dox	M	16	4	1	NA	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	No
B536	NL (Protocol 3) + dox	M	10	3	0	NA	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	No
B538	NL (Protocol 3) + dox	M	10	3	0	1	IGH/MYC	NA	Yes	Yes	2	No	Yes	No	Yes	No

B550	NL (Protocol 3) + dox	F	5	3 or 4	0	0	MYC/IGL	NA	Yes	Yes	2	No	No	No	No	No
B555	NA	NA	NA	NA	NA	NA	IGH/MYC	NA	No	No	2	No	No	No	No	No
B558	NL (Protocol 3) + dox	F	7.5	3	0	0	IGH/MYC	NA	Yes	Yes	2	No	No	Yes	No	No
B570	NL (Protocol 3) + dox	F	12	3	0	0	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	Yes
B572	NL (Protocol 3) + dox	F	7	3 or 4	NA	NA	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	No
B583	NL (Protocol 3) + dox	F	7	4	NA	1	IGH/MYC	NA	Yes	Yes	2	No	No	Yes	Yes	No
B595	NL (Protocol 3) + dox	F	6	1	0	0	MYC/IGL	NA	Yes	Yes	2	No	No	Yes	Yes	No
B618	NL (Protocol 3) + dox	F	5	3	0	0	MYC/IGL	NA	Yes	Yes	2	No	No	No	No	No
B624	NL (Protocol 3) + dox	F	8	1	0	0	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	No
B631	NL (Protocol 3) + dox	F	7	3	0	0	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	No
B643	NL (Protocol 3) + dox	F	12	4	1	0	IGH/MYC	NA	Yes	Yes	2	No	Yes	Yes	No	No
B648	NL (Protocol 3) + dox	M	8	3	0	0	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	No
B667	NL (Protocol 3) + dox	M	7	3 or 4	0	0	MYC/IGL	NA	Yes	Yes	2	No	No	No	No	No
B674	NL (Protocol 3) + dox	M	7	3	0	0	IGH/MYC	NA	Yes	Yes	2	No	Yes	No	Yes	No
B676	NL (Protocol 3) + dox	F	7	3	0	0	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	No
B678	NL (Protocol 3) + dox	F	9	4	0	0	MYC/IGL	NA	Yes	Yes	2	No	No	No	No	No
B679	NL (Protocol 3) + dox	M	11	4	0	1	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	No
B680	NL (Protocol 3) + dox	M	9	4	NA	0	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	No
B681	NL (Protocol 3) + dox	F	12	4	0	1	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	Yes
B686	NL (Protocol 3) + dox	M	3	3	0	0	IGH/MYC	NA	Yes	Yes	2	No	No	Yes	No	No
B697	NL (Protocol 3) + dox	F	9	3	0	0	IGH/MYC	NA	Yes	Yes	2	No	Yes	No	No	Yes
B709	NL (Protocol 3) + dox	M	4	3	0	NA	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	No
B721	NL (Protocol 3) + dox	M	6	4	0	1	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	No
B730	NL (Protocol 3) + dox	M	9	3	0	NA	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	No
B738	NL (Protocol 3) + dox	M	3	3	0	0	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	Yes

B755	NL (Protocol 3) + dox	M	8	3	0	0	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	No
B766	NL (Protocol 3) + dox	F	8	3	0	0	IGH/MYC	NA	Yes	Yes	2	No	No	Yes	No	No
B767	NA	F	5	3 or 4	NA	NA	IGH/MYC	NA	No	No	2	Yes	Yes	No	Yes	Yes
B773	NL (Protocol 3) + dox	M	14	3	0	0	MYC/IGL	NA	Yes	Yes	2	No	No	No	No	No
B779	NL (Protocol 3) + dox	M	13	1	0	0	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	No
B780	NL (Protocol 3) + dox	M	8	3	0	0	MYC/IGL	NA	Yes	Yes	2	No	Yes	Yes	Yes	No
B800	NL (Protocol 3) + dox	F	4	1	0	0	IGH/MYC	NA	Yes	Yes	2	No	No	No	No	No
B803	NL (Protocol 3) + dox	M	10	4	0	1	MYC/IGL	NA	Yes	Yes	2	No	No	No	Yes	Yes
B808	NL (Protocol 3) + dox	M	5	1	0	0	MYC/IGL	NA	Yes	Yes	2	No	No	Yes	No	No

Appendix G - Clinical and Genomic Data for 105 endemic BL cases.

Appendix H

B number	17p Abnormality	*TP53* Mutation	Nucleotide	Polyphen	Condel	*TP53* Status	*TP53* abnormality	Note
B095	No abnormality	None				Normal	No	
B105	No abnormality	None				Normal	No	
B121	No abnormality	None				Normal	No	
B150	No abnormality	None				Normal	No	
B153	No abnormality	None				Normal	No	
B156	No abnormality	None				Normal	No	
B160	17p deletion	None				Monoallelic	Yes	Loss + No mutation
B172	17p deletion	None				Monoallelic	Yes	Loss + No mutation
B174	17p deletion	None				Monoallelic	Yes	Loss + No mutation
B186	17p deletion	None				Monoallelic	Yes	Loss + No mutation
B190	No abnormality	**p.P177S**	c.529C>T	Probably damaging	Deleterious	Monoallelic	Yes	1 Mutation only
B195	No abnormality	None				Normal	No	
B201	No abnormality	None				Normal	No	
B210	No abnormality	None				Normal	No	
B216	No abnormality	None				Normal	No	
B232	17p deletion	**p.R248W**	c.742C>T	Probably damaging	Deleterious	Biallelic	Yes	Hemizygous
B264	No abnormality	None				Normal	No	
B272	No abnormality	None				Normal	No	
B273	No abnormality	None				Normal	No	
B282	17p CNN-LOH	**p.Y234C**	c.701A>G	Probably damaging	Deleterious	Biallelic	Yes	Homozygous
B284	No abnormality	None				Normal	No	
B285	17p deletion	None				Monoallelic	Yes	Loss + No mutation

B286	No abnormality	**p.R283P**	c.848G>C	Probably damaging	Deleterious	Monoallelic	Yes	1 Mutation only
B287	No abnormality	None				Normal	No	
B288	No abnormality	**p.L257Q**	c.770T>A	Probably damaging	Deleterious	Monoallelic	Yes	1 Mutation only
B289	No abnormality	**p.R280G**	c.838A>G	Probably damaging	Deleterious	Monoallelic	Yes	1 Mutation only
B294	No abnormality	None				Normal	No	
B298	No abnormality	**p.Y234N**	c.700T>A	Probably damaging	Deleterious	Monoallelic	Yes	1 Mutation only
B309	No abnormality	None				Normal	No	
B311	No abnormality	None				Normal	No	
B313	No abnormality	None				Normal	No	
B314	No abnormality	None				Normal	No	
B321	No abnormality	None				Normal	No	
B324	No abnormality	None				Normal	No	
B326	No abnormality	None				Normal	No	
B330	No abnormality	**p.R273C, p.Y205H**	c.817C>T, c.613T>C	Probably damaging, Probably damaging	Deleterious, Neutral	Biallelic	Yes	2 Mutations
B343	No abnormality	None				Normal	No	
B346	No abnormality	**p.R213Q**	c.638G>A	Probably damaging	Deleterious	Monoallelic	Yes	1 Mutation only
B347	No abnormality	**p.C135W**	c.405C>G	Probably damaging	Deleterious	Monoallelic	Yes	1 Mutation only
B349	No abnormality	None				Normal	No	
B361	No abnormality	None				Normal	No	
B363	17p deletion	None				Monoallelic	Yes	Loss + No mutation
B370	No abnormality	**p.Y236N**	c.706T>A	Probably damaging	Deleterious	Monoallelic	Yes	1 Mutation only
B371	No abnormality	None				Normal	No	
B374	No abnormality	**p.Y234R**	c.700_701delTAinsCG	Possibly damaging	No Data	Monoallelic	Yes	1 Mutation only
B375	No abnormality	None				Normal	No	
B389	No abnormality	None				Normal	No	

B391	No abnormality	**p.S240R**	c.718A>C	Probably damaging	Deleterious	Monoallelic	Yes	1 Mutation only
B397	No abnormality	**p.R175C**	c.523C>T	Probably damaging	Deleterious	Monoallelic	Yes	1 Mutation only
B409	No abnormality	**p.R175H**	c.524G>A	Possibly damaging	Deleterious	Monoallelic	Yes	1 Mutation only
B410	No abnormality	**p.R248W**	c.742C>T	Probably damaging	Deleterious	Monoallelic	Yes	1 Mutation only
B413	No abnormality	**p.G245D**	c.734G>A	Probably damaging	Deleterious	Monoallelic	Yes	1 Mutation only
B420	No abnormality	None				Normal	No	
B422	17p deletion	**p.R248Q**	c.743G>A	Probably damaging	Deleterious	Biallelic	Yes	Hemizygous
B423	No abnormality	None				Normal	No	
B428	No abnormality	**p.I232T**	c.695T>C	Benign	Neutral	Monoallelic	Yes	1 Mutation only
B436	No abnormality	**p.T231A, p.R249G**	c.691A>G, c.745A>G	Benign, Probably damaging	Neutral, Deleterious	Biallelic	Yes	2 Mutations
B453	17p CNN-LOH	**p.Y205F**	c.613_614ins(1)	No data (Frameshift)	No data (Frameshift)	Biallelic	Yes	Homozygous
B459	No abnormality	None				Normal	No	
B463	No abnormality	**p.M246V**	c.736A>G	Probably damaging	Neutral	Monoallelic	Yes	1 Mutation only
B483	No abnormality	NA				NA		
B492	No abnormality	None				Normal	No	
B495	No abnormality	**p.H179L**	c.536A>T	Benign	Deleterious	Monoallelic	Yes	1 Mutation only
B498	No abnormality	NA				NA		
B501	No abnormality	None				Normal	No	
B505	No abnormality	**p.Y234R**	c.700_701delTAinsCG	Possibly damaging	No Data	Monoallelic	Yes	1 Mutation only
B509	No abnormality	None				Normal	No	
B516	No abnormality	NA				NA		
B520	No abnormality	None				Normal	No	
B536	No abnormality	None				Normal	No	
B538	No abnormality	**p.R175H**	c.524G>A	Possibly damaging	Deleterious	Monoallelic	Yes	1 Mutation only
B550	No abnormality	None				Normal	No	

B555	No abnormality	None				Normal	No	
B558	17p deletion	**p.R248W**	c.742C>T	Probably damaging	Deleterious	Biallelic	Yes	Hemizygous
B570	No abnormality	None				Normal	No	
B572	No abnormality	**p.C182***	c.546C>A	No data (nonsense)	No data (nonsense)	Monoallelic	Yes	1 Mutation only
B583	17p deletion	None				Monoallelic	Yes	Loss + No mutation
B595	17p deletion	None				Monoallelic	Yes	Loss + No mutation
B618	No abnormality	**p.P190L**	c.569C>T	Possibly damaging	Neutral	Monoallelic	Yes	1 Mutation only
B624	No abnormality	None				Normal	No	
B631	No abnormality	**p.P151L**	c.569C>T	Possibly damaging	Neutral	Monoallelic	Yes	1 Mutation only
B643	17p deletion	None				Monoallelic	Yes	Loss + No mutation
B648	No abnormality	None				Normal	No	
B667	No abnormality	**p.K132R**	c.395A>G	Probably damaging	Deleterious	Monoallelic	Yes	1 Mutation only
B674	No abnormality	None				Normal	No	
B676	No abnormality	**p.P152L**	c.455C>T	Probably damaging	Deleterious	Monoallelic	Yes	1 Mutation only
B678	No abnormality	None				Normal	No	
B679	No abnormality	None				Normal	No	
B680	17p CNN-LOH	**p.R282W**	c.844C>T	Probably damaging	Deleterious	Biallelic	Yes	Homozygous
B681	No abnormality	None				Normal	No	
B686	17p deletion	**p.R175H**	c.524G>A	Possibly damaging	Deleterious	Biallelic	Yes	Hemizygous
B697	No abnormality	None				Normal	No	
B709	No abnormality	None				Normal	No	
B721	No abnormality	None				Normal	No	
B730	No abnormality	NA				NA		
B738	No abnormality	None				Normal	No	
B755	No abnormality	None				Normal	No	

B766	17p deletion	None				Monoallelic	Yes	Loss + No mutation
B767	17p CNN-LOH	**p.Y236C**	c.707A>G	Probably damaging	Deleterious	Biallelic	Yes	Homozygous
B773	No abnormality	None				Normal	No	
B779	No abnormality	**p.R273C**	c.817C>T	Probably damaging	Deleterious	Monoallelic	Yes	1 Mutation only
B780	17p deletion	**p.C176S**	c.526T>A	Possibly damaging	Deleterious	Biallelic	Yes	Hemizygous
B800	No abnormality	None				Normal	No	
B803	No abnormality	None				Normal	No	
B808	17p deletion	**p.Y234C, p.M237I**	c.701A>G, c.711G>A	Probably damaging, Probably damaging	Deleterious, Deleterious	Biallelic	Yes	Hemizygous

Appendix H - Detailed breakdown of TP53 aberrations in endemic BL cases.

Appendix I

Publication entitled “Sporadic and endemic Burkitt lymphoma have frequent *FOXO1* mutations but distinct hotspots in the AKT recognition motif” published in *Blood Advances*. Attached to end of document.

Appendix J

Manuscript entitled “Genomic abnormalities of *TP53* define a high-risk subgroup of paediatric B-cell non-Hodgkin lymphoma” submitted to the *Journal of Clinical Oncology* and returned with reviewers’ comments for re-submission in March 2020. Attached to end of document.

Bibliography

Abate, F., Ambrosio, M.R., Mundo, L., Laginestra, M.A., Fuligni, F., Rossi, M., Zairis, S., Gazaneo, S., De Falco, G., Lazzi, S., Bellan, C., Rocca, B.J., Amato, T., Marasco, E., Etebari, M., Ogwang, M., Calbi, V., Ndede, I., Patel, K., Chumba, D., Piccaluga, P.P., Pileri, S., Leoncini, L. and Rabadan, R. (2015) 'Distinct Viral and Mutational Spectrum of Endemic Burkitt Lymphoma', *PLoS Pathog*, 11(10), p. e1005158.

Abe, H., Maeda, D., Hino, R., Otake, Y., Isogai, M., Ushiku, A.S., Matsusaka, K., Kunita, A., Ushiku, T., Uozaki, H., Tateishi, Y., Hishima, T., Iwasaki, Y., Ishikawa, S. and Fukayama, M. (2012) 'ARID1A expression loss in gastric cancer: pathway-dependent roles with and without Epstein-Barr virus infection and microsatellite instability', *Virchows Arch*, 461(4), pp. 367-77.

Adzhubei, I., Jordan, D.M. and Sunyaev, S.R. (2013) 'Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2', *Curr Protoc Hum Genet*, 0 7, p. Unit7 20.

Agrawal, A., Yang, J., Murphy, R.F. and Agrawal, D.K. (2006) 'Regulation of the p14ARF-Mdm2-p53 pathway: an overview in breast cancer', *Exp Mol Pathol*, 81(2), pp. 115-22.

Alhejaily, A., Day, A.G., Feilotter, H.E., Baetz, T. and Lebrun, D.P. (2014) 'Inactivation of the CDKN2A tumor-suppressor gene by deletion or methylation is common at diagnosis in follicular lymphoma and associated with poor clinical outcome', *Clin Cancer Res*, 20(6), pp. 1676-86.

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., Powell, J.I., Yang, L., Marti, G.E., Moore, T., Hudson, J., Jr., Lu, L., Lewis, D.B., Tibshirani, R., Sherlock, G., Chan, W.C., Greiner, T.C., Weisenburger, D.D., Armitage, J.O., Warnke, R., Levy, R., Wilson, W., Grever, M.R., Byrd, J.C., Botstein, D., Brown, P.O. and Staudt, L.M. (2000) 'Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling', *Nature*, 403(6769), pp. 503-11.

Andrews, S. (2010) *FastQC: a quality control tool for high throughput sequence data.* [Computer program]. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Anoop, P., Sankpal, S., Stiller, C., Tewari, S., Lancaster, D.L., Khabra, K. and Taj, M.M. (2012) 'Outcome of childhood relapsed or refractory mature B-cell non-Hodgkin lymphoma and acute lymphoblastic leukemia', *Leuk Lymphoma*, 53(10), pp. 1882-8.

Asic, K. (2016) 'Dominant mechanisms of primary resistance differ from dominant mechanisms of secondary resistance to targeted therapies', *Crit Rev Oncol Hematol*, 97, pp. 178-96.

Aukema, S.M., Theil, L., Rohde, M., Bauer, B., Bradtke, J., Burkhardt, B., Bonn, B.R., Claviez, A., Gattenlohner, S., Makarova, O., Nagel, I., Oschlies, I., Pott, C., Szczepanowski, M., Traulsen, A., Kluin, P.M., Klapper, W., Siebert, R. and Murga Penas, E.M. (2015) 'Sequential karyotyping in Burkitt lymphoma reveals a linear clonal evolution with increase in karyotype complexity and a high frequency of recurrent secondary aberrations', *Br J Haematol*, 170(6), pp. 814-25.

Banh, R.S., Iorio, C., Marcotte, R., Xu, Y., Cojocari, D., Rahman, A.A., Pawling, J., Zhang, W., Sinha, A., Rose, C.M., Isasa, M., Zhang, S., Wu, R., Virtanen, C., Hitomi, T., Habu, T., Sidhu, S.S., Koizumi, A., Wilkins, S.E., Kislinger, T., Gygi, S.P., Schofield, C.J., Dennis, J.W., Wouters, B.G. and Neel, B.G. (2016) 'PTP1B controls non-mitochondrial oxygen consumption by regulating RNF213 to promote tumour survival during hypoxia', *Nat Cell Biol*, 18(7), pp. 803-813.

Basso, K. and Dalla-Favera, R. (2015) 'Germinal centres and B cell lymphomagenesis', *Nat Rev Immunol*, 15(3), pp. 172-84.

Baugh, E.H., Ke, H., Levine, A.J., Bonneau, R.A. and Chan, C.S. (2018) 'Why are there hotspot mutations in the TP53 gene in human cancers?', *Cell Death Differ*, 25(1), pp. 154-160.

Bendjennat, M., Boulaire, J., Jascur, T., Brickner, H., Barbier, V., Sarasin, A., Fotedar, A. and Fotedar, R. (2003) 'UV irradiation triggers ubiquitin-dependent degradation of p21(WAF1) to promote DNA repair', *Cell*, 114(5), pp. 599-610.

Benjamin, D.S., T. Cibulskis, K. Getz, G. Stewart, C. Lichtenstein, L. (2019) 'Calling Somatic SNVs and Indels with Mutect2', *bioRxiv*.

Beroukhim, R., Getz, G., Nghiemphu, L., Barretina, J., Hsueh, T., Linhart, D., Vivanco, I., Lee, J.C., Huang, J.H., Alexander, S., Du, J., Kau, T., Thomas, R.K., Shah, K., Soto, H., Perner, S.,

Prensner, J., Debiasi, R.M., Demichelis, F., Hatton, C., Rubin, M.A., Garraway, L.A., Nelson, S.F., Liau, L., Mischel, P.S., Cloughesy, T.F., Meyerson, M., Golub, T.A., Lander, E.S., Mellinghoff, I.K. and Sellers, W.R. (2007) 'Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma', *Proc Natl Acad Sci U S A*, 104(50), pp. 20007-12.

Blombery, Wall, M. and Seymour, J. (2015) 'The molecular pathogenesis of B-cell non-Hodgkin lymphoma', *European journal of haematology*, 95(4).

Bouaoun, L., Sonkin, D., Ardin, M., Hollstein, M., Byrnes, G., Zavadil, J. and Olivier, M. (2016) 'TP53 Variations in Human Cancers: New Lessons from the IARC TP53 Database and Genomics Data', *Hum Mutat*, 37(9), pp. 865-76.

Brown, J.R., Byrd, J.C., Coutre, S.E., Benson, D.M., Flinn, I.W., Wagner-Johnston, N.D., Spurgeon, S.E., Kahl, B.S., Bello, C., Webb, H.K., Johnson, D.M., Peterman, S., Li, D., Jahn, T.M., Lannutti, B.J., Ulrich, R.G., Yu, A.S., Miller, L.L. and Furman, R.R. (2014) 'Idelalisib, an inhibitor of phosphatidylinositol 3-kinase p110δ, for relapsed/refractory chronic lymphocytic leukemia', in *Blood*. pp. 3390-7.

Burkitt, D. (1958) 'A sarcoma involving the jaws in African children', *Br J Surg*, 46(197), pp. 218-23.

Burmeister, T., Molkentin, M., Schwartz, S., Gokbuget, N., Hoelzer, D., Thiel, E. and Reinhardt, R. (2013) 'Erroneous class switching and false VDJ recombination: molecular dissection of t(8;14)/MYC-IGH translocations in Burkitt-type lymphoblastic leukemia/B-cell lymphoma', *Mol Oncol*, 7(4), pp. 850-8.

Béguelin, W., Popovic, R., Teater, M., Jiang, Y., Bunting, K.L., Rosen, M., Shen, H., Yang, S.N., Wang, L., Ezponda, T., Martinez-Garcia, E., Zhang, H., Zhang, Y., Verma, S.K., McCabe, M.T., Ott, H.M., Van Aller, G.S., Kruger, R.G., Liu, Y., McHugh, C.F., Scott, D.W., Chung, Y.R., Kelleher, N., Shaknovich, R., Creasy, C.L., Gascoyne, R.D., Wong, K.K., Cerchietti, L.C., Levine, R.L., Abdel-Wahab, O., Licht, J.D., Elemento, O. and Melnick, A.M. (2013) 'EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation', *Cancer Cell*, 23(5), pp. 677-92.

Cairo, M., Auperin, A., Perkins, S.L., Pinkerton, R., Harrison, L., Goldman, S. and Patte, C. (2018) 'Overall survival of children and adolescents with mature B cell non-Hodgkin lymphoma who had refractory or relapsed disease during or after treatment with FAB/LMB 96: A report from the FAB/LMB 96 study group', *Br J Haematol*, 182(6), pp. 859-869.

Calado, D.P., Sasaki, Y., Godinho, S.A., Pellerin, A., Köchert, K., Sleckman, B.P., de Alborán, I.M., Janz, M., Rodig, S. and Rajewsky, K. (2012) 'MYC is essential for the formation and maintenance of germinal centers', *Nat Immunol*, 13(11), pp. 1092-100.

Campbell, K.J., Bath, M.L., Turner, M.L., Vandenberg, C.J., Bouillet, P., Metcalf, D., Scott, C.L. and Cory, S. (2010) 'Elevated Mcl-1 perturbs lymphopoiesis, promotes transformation of hematopoietic stem/progenitor cells, and enhances drug resistance', *Blood*, 116(17), pp. 3197-207.

Campo, E., Cymbalista, F., Ghia, P., Jäger, U., Pospisilova, S., Rosenquist, R., Schuh, A. and Stilgenbauer, S. (2018) 'TP53 aberrations in chronic lymphocytic leukemia: an overview of the clinical implications of improved diagnostics', in *Haematologica*. pp. 1956-68.

Capoulade, B, B.-d.P., I, L., M, R., J, F., T, T. and J, W. (1998) 'Overexpression of MDM2, due to enhanced translation, results in inactivation of wild-type p53 in Burkitt's lymphoma cells', *Oncogene*, 16(12).

Castellino, R.C., De Bortoli, M., Lin, L.L., Skapura, D.G., Rajan, J.A., Adesina, A.M., Perlaky, L., Irwin, M.S. and Kim, J.Y. (2007) 'Overexpressed TP73 induces apoptosis in medulloblastoma', *BMC Cancer*, 7, p. 127.

Challa-Malladi, M., Lieu, Y.K., Califano, O., Holmes, A., Bhagat, G., Murty, V.V., Dominguez-Sola, D., Pasqualucci, L. and Dalla-Favera, R. (2011) 'Combined Genetic Inactivation of Beta2-Microglobulin and CD58 Reveals Frequent Escape from Immune Recognition in Diffuse Large B-cell Lymphoma', *Cancer Cell*, 20(6), pp. 728-40.

Chang, C.C., Liu, Y.C., Cleveland, R.P. and Perkins, S.L. (2000) 'Expression of c-Myc and p53 correlates with clinical outcome in diffuse large B-cell lymphomas', *Am J Clin Pathol*, 113(4), pp. 512-8.

Chen, J., Lui, W.O., Vos, M.D., Clark, G.J., Takahashi, M., Schoumans, J., Khoo, S.K., Petillo, D., Lavery, T., Sugimura, J., Astuti, D., Zhang, C., Kagawa, S., Maher, E.R., Larsson, C., Alberts,

A.S., Kanayama, H.O. and Teh, B.T. (2003) 'The t(1;3) breakpoint-spanning genes LSAMP and NORE1 are involved in clear cell renal cell carcinomas', *Cancer Cell*, 4(5), pp. 405-13.

Cho-Vega, J.H., Rassidakis, G.Z., Admirand, J.H., Oyarzo, M., Ramalingam, P., Paraguya, A., McDonnell, T.J., Amin, H.M. and Medeiros, L.J. (2004) 'MCL-1 expression in B-cell non-Hodgkin's lymphomas', *Hum Pathol*, 35(9), pp. 1095-100.

Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S. and Getz, G. (2013) 'Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples', *Nat Biotechnol*, 31(3), pp. 213-9.

Cimmino, A., Calin, G.A., Fabbri, M., Iorio, M.V., Ferracin, M., Shimizu, M., Wojcik, S.E., Aqeilan, R.I., Zupo, S., Dono, M., Rassenti, L., Alder, H., Volinia, S., Liu, C.G., Kipps, T.J., Negrini, M. and Croce, C.M. (2005) 'miR-15 and miR-16 induce apoptosis by targeting BCL2', *Proc Natl Acad Sci U S A*, 102(39), pp. 13944-9.

Conde, L., Riby, J., Zhang, J., Bracci, P.M. and Skibola, C.F. (2014) 'Copy number variation analysis on a non-Hodgkin lymphoma case-control study identifies an 11q25 duplication associated with diffuse large B-cell lymphoma', *PLoS One*, 9(8), p. e105382.

Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M.R., Armean, I.M., Bennett, R., Bhai, J., Billis, K., Boddu, S., Cummins, C., Davidson, C., Dodiya, K.J., Gall, A., Giron, C.G., Gil, L., Grego, T., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O.G., Janacek, S.H., Juettemann, T., Kay, M., Laird, M.R., Lavidas, I., Liu, Z., Loveland, J.E., Marugan, J.C., Maurel, T., McMahon, A.C., Moore, B., Morales, J., Mudge, J.M., Nuhn, M., Ogeh, D., Parker, A., Parton, A., Patricio, M., Abdul Salam, A.I., Schmitt, B.M., Schuilenburg, H., Sheppard, D., Sparrow, H., Stapleton, E., Szuba, M., Taylor, K., Threadgold, G., Thormann, A., Vullo, A., Walts, B., Winterbottom, A., Zadissa, A., Chakiachvili, M., Frankish, A., Hunt, S.E., Kostadima, M., Langridge, N., Martin, F.J., Muffato, M., Perry, E., Ruffier, M., Staines, D.M., Trevanion, S.J., Aken, B.L., Yates, A.D., Zerbino, D.R. and Flicek, P. (2019) 'Ensembl 2019', *Nucleic Acids Res*, 47(D1), pp. D745-d751.

Dal Bo, M., Bomben, R., Hernández, L. and Gattei, V. (2015) 'The MYC/miR-17-92 axis in lymphoproliferative disorders: A common pathway with therapeutic potential', *Oncotarget*, 6(23), pp. 19381-92.

Dalla-Favera, R., Bregni, M., Erikson, J., Patterson, D., Gallo, R.C. and Croce, C.M. (1982) 'Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells', *Proc Natl Acad Sci U S A*, 79(24), pp. 7824-7.

Dang, C.V. (2012) 'MYC on the Path to Cancer', *Cell*, 149(1), pp. 22-35.

Dave, S.S., Fu, K., Wright, G.W., Lam, L.T., Kluin, P., Boerma, E.J., Greiner, T.C., Weisenburger, D.D., Rosenwald, A., Ott, G., Muller-Hermelink, H.K., Gascoyne, R.D., Delabie, J., Rimsza, L.M., Braziel, R.M., Grogan, T.M., Campo, E., Jaffe, E.S., Dave, B.J., Sanger, W., Bast, M., Vose, J.M., Armitage, J.O., Connors, J.M., Smeland, E.B., Kvaloy, S., Holte, H., Fisher, R.I., Miller, T.P., Montserrat, E., Wilson, W.H., Bahl, M., Zhao, H., Yang, L., Powell, J., Simon, R., Chan, W.C. and Staudt, L.M. (2006) 'Molecular diagnosis of Burkitt's lymphoma', *N Engl J Med*, 354(23), pp. 2431-42.

de Fontbrune, F., Canioni, D., Chapdelaine, H., Delarue, R., Brousse, N., Suarez, F., Hermine, O. (2019) 'High Level of CD71 Expression On Neoplastic B Cells At Diagnosis Is Predictive of Overall Survival After Rituximab and Anthracyclines Based Regimen in Follicular Lymphoma | Blood | American Society of Hematology', *Blood*, 120(21).

de Leval, L. and Harris, N.L. (2003) 'Variability in immunophenotype in diffuse large B-cell lymphoma and its clinical relevance', *Histopathology*, 43(6), pp. 509-28.

Deffenbacher, K.E., Iqbal, J., Sanger, W., Shen, Y., Lachel, C., Liu, Z., Liu, Y., Lim, M.S., Perkins, S.L., Fu, K., Smith, L., Lynch, J., Staudt, L.M., Rimsza, L.M., Jaffe, E., Rosenwald, A., Ott, G.K., Delabie, J., Campo, E., Gascoyne, R.D., Cairo, M.S., Weisenburger, D.D., Greiner, T.C., Gross, T.G. and Chan, W.C. (2012) 'Molecular distinctions between pediatric and adult mature B-cell non-Hodgkin lymphomas identified through genomic profiling', *Blood*, 119(16), pp. 3757-66.

Dent, A.L., Shaffer, A.L., Yu, X., Allman, D. and Staudt, L.M. (1997) 'Control of inflammation, cytokine expression, and germinal center formation by BCL-6', *Science*, 276(5312), pp. 589-92.

Dixit, V.M., Green, S., Sarma, V., Holzman, L.B., Wolf, F.W., O'Rourke, K., Ward, P.A., Prochownik, E.V. and Marks, R.M. (1990) 'Tumor necrosis factor-alpha induction of novel gene products in human endothelial cells including a macrophage-specific chemotaxin', *J Biol Chem*, 265(5), pp. 2973-8.

Dominguez-Sola, GD, V., CY, Y., RT, P., M, S., MC, N. and R, D.-F. (2012) 'The proto-oncogene MYC is required for selection in the germinal center and cyclic reentry', *Nature immunology*, 13(11).

Donehower, L.A., Soussi, T., Korkut, A., Liu, Y., Schultz, A., Cardenas, M., Li, X., Babur, O., Hsu, T.K., Lichtarge, O., Weinstein, J.N., Akbani, R., Cancer Genome Atlas, N. and Wheeler, D.A. (2019) 'Integrated Analysis of TP53 Gene and Pathway Alterations in The Cancer Genome Atlas', *Cell Rep*, 28(5), pp. 1370-1384 e5.

Duan, S., Cermak, L., Pagan, J.K., Rossi, M., Martinengo, C., di Celle, P.F., Chapuy, B., Shipp, M., Chiarle, R. and Pagano, M. (2012) 'FBXO11 targets BCL6 for degradation and is inactivated in diffuse large B-cell lymphomas', *Nature*, 481(7379), pp. 90-3.

Duesberg, P.H. and Vogt, P.K. (1979) 'Avian acute leukemia viruses MC29 and MH2 share specific RNA sequences: evidence for a second class of transforming genes', *Proc Natl Acad Sci U S A*, 76(4), pp. 1633-7.

Edlinger, L., Berger-Becvar, A., Menzl, I., Hoermann, G., Greiner, G., Grundschober, E., Bago-Horvath, Z., Al-Zoughbi, W., Hoefler, G., Brostjan, C., Gille, L., Moriggl, R., Spittler, A., Sexl, V. and Hoelbl-Kovacic, A. (2017) 'Expansion of BCR/ABL1(+) cells requires PAK2 but not PAK1', *Br J Haematol*, 179(2), pp. 229-241.

Eischen, C.M., Weber, J.D., Roussel, M.F., Sherr, C.J. and Cleveland, J.L. (1999) 'Disruption of the ARF–Mdm2–p53 tumor suppressor pathway in Myc-induced lymphomagenesis', *Genes Dev*, 13(20), pp. 2658-69.

Ellis, L.M. and Hicklin, D.J. (2009) 'Resistance to Targeted Therapies: Refining Anticancer Therapy in the Era of Molecular Oncology', *Clin Cancer Res*, 15(24), pp. 7471-7478.

Epstein, M.A., Achong, B.G. and Barr, Y.M. (1964) 'VIRUS PARTICLES IN CULTURED LYMPHOBLASTS FROM BURKITT'S LYMPHOMA', *Lancet*, 1(7335), pp. 702-3.

Escudero-Esparza, A., Bartoschek, M., Gialeli, C., Okroj, M., Owen, S., Jirstrom, K., Orimo, A., Jiang, W.G., Pietras, K. and Blom, A.M. (2016) 'Complement inhibitor CSMD1 acts as tumor suppressor in human breast cancer', *Oncotarget*, 7(47), pp. 76920-76933.

Ferreiro, J.F., Morscio, J., Dierickx, D., Marcelis, L., Verhoef, G., Vandenberghe, P., Tousseyn, T. and Wlodarska, I. (2015) 'Post-transplant molecularly defined Burkitt lymphomas are frequently MYC-negative and characterized by the 11q-gain/loss pattern', *Haematologica*, 100(7), pp. e275-9.

Ferreri, A.J., Govi, S., Pileri, S.A. and Savage, K.J. (2012) 'Anaplastic large cell lymphoma, ALK-positive', *Crit Rev Oncol Hematol*, 83(2), pp. 293-302.

Ferry, J.A. (2006) 'Burkitt's lymphoma: clinicopathologic features and differential diagnosis', *Oncologist*, 11(4), pp. 375-83.

Fitzsimmons, L. and Kelly, G.L. (2017) 'EBV and Apoptosis: The Viral Master Regulator of Cell Fate?', *Viruses*, 9(11).

Fujita, N., Mori, T., Mitsui, T., Inada, H., Horibe, K. and Tsurusawa, M. (2008) 'The role of hematopoietic stem cell transplantation with relapsed or primary refractory childhood B-cell non-Hodgkin lymphoma and mature B-cell leukemia: a retrospective analysis of enrolled cases in Japan', *Pediatr Blood Cancer*, 51(2), pp. 188-92.

Gallagher, C.J., Kadlubar, F.F., Muscat, J.E., Ambrosone, C.B., Lang, N.P. and Lazarus, P. (2007) 'The UGT2B17 gene deletion polymorphism and risk of prostate cancer. A case-control study in Caucasians', *Cancer Detect Prev*, 31(4), pp. 310-5.

Gao, Y., Wu, H., He, D., Hu, X. and Li, Y. (2013) 'Downregulation of BCL11A by siRNA induces apoptosis in B lymphoma cell lines', in *Biomed Rep*. pp. 47-52.

Gilmore, T.D. and Gerondakis, S. (2011) 'The c-Rel Transcription Factor in Development and Disease', in *Genes Cancer*. pp. 695-711.

Gindin, T., Murty, V., Alobeid, B. and Bhagat, G. (2015) 'MLL/KMT2A translocations in diffuse large B-cell lymphomas', *Hematol Oncol*, 33(4), pp. 239-46.

Gisselbrecht, C., Glass, B., Mounier, N., Singh Gill, D., Linch, D.C., Trneny, M., Bosly, A., Ketterer, N., Shpilberg, O., Hagberg, H., Ma, D., Briere, J., Moskowitz, C.H. and Schmitz, N. (2010) 'Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in the rituximab era', *J Clin Oncol*, 28(27), pp. 4184-90.

Giulino-Roth, L. and Goldman, S. (2016) 'Recent molecular and therapeutic advances in B-cell non-Hodgkin lymphoma in children', *Br J Haematol*, 173(4), pp. 531-44.

Giulino-Roth, L., Wang, K., MacDonald, T.Y., Mathew, S., Tam, Y., Cronin, M.T., Palmer, G., Lucena-Silva, N., Pedrosa, F., Pedrosa, M., Teruya-Feldstein, J., Bhagat, G., Alobeid, B., Leoncini, L., Bellan, C., Rogena, E., Pinkney, K.A., Rubin, M.A., Ribeiro, R.C., Yelensky, R., Tam, W., Stephens, P.J. and Cesarman, E. (2012) 'Targeted genomic sequencing of pediatric Burkitt lymphoma identifies recurrent alterations in antiapoptotic and chromatin-remodeling genes', *Blood*, 120(26), pp. 5181-4.

Gonzalez-Farre, B., Ramis-Zaldivar, J.E., Salmeron-Villalobos, J., Balague, O., Celis, V., Verdu-Amoros, J., Nadeu, F., Sabado, C., Ferrandez, A., Garrido, M., Garcia-Bragado, F., de la Maya, M.D., Vagace, J.M., Panizo, C.M., Astigarraga, I., Andres, M., Jaffe, E.S., Campo, E. and Salaverria, I. (2019) 'Burkitt-like lymphoma with 11q aberration: a germinal center-derived lymphoma genetically unrelated to Burkitt lymphoma', *Haematologica*, 104(9), pp. 1822-1829.

Grande, B.M., Gerhard, D.S., Jiang, A., Griner, N.B., Abramson, J.S., Alexander, T.B., Allen, H., Ayers, L.W., Bethony, J.M., Bhatia, K., Bowen, J., Casper, C., Choi, J.K., Culibrk, L., Davidsen, T.M., Dyer, M.A., Gastier-Foster, J.M., Gesuwan, P., Greiner, T.C., Gross, T.G., Hanf, B., Harris, N.L., He, Y., Irvin, J.D., Jaffe, E.S., Jones, S.J.M., Kerchan, P., Knoetze, N., Leal, F.E., Lichtenberg, T.M., Ma, Y., Martin, J.P., Martin, M.R., Mbulaiteye, S.M., Mullighan, C.G., Mungall, A.J., Namirembe, C., Novik, K., Noy, A., Ogwang, M.D., Omoding, A., Orem, J., Reynolds, S.J., Rushton, C.K., Sandlund, J.T., Schmitz, R., Taylor, C., Wilson, W.H., Wright, G.W., Zhao, E.Y., Marra, M.A., Morin, R.D. and Staudt, L.M. (2019) 'Genome-wide discovery of somatic coding and noncoding mutations in pediatric endemic and sporadic Burkitt lymphoma', *Blood*, 133(12), pp. 1313-1324.

Grumont, R.J., Rourke, I.J., O'Reilly, L.A., Strasser, A., Miyake, K., Sha, W. and Gerondakis, S. (1998) 'B Lymphocytes Differentially Use the Rel and Nuclear Factor κB1 (NF-κB1) Transcription Factors to Regulate Cell Cycle Progression and Apoptosis in Quiescent and Mitogen-activated Cells', *J Exp Med*, 187(5), pp. 663-74.

Gyorffy, B., Pongor, L., Bottai, G., Li, X., Budczies, J., Szabo, A., Hatzis, C., Pusztai, L. and Santarpia, L. (2018) 'An integrative bioinformatics approach reveals coding and non-coding

gene variants associated with gene expression profiles and outcome in breast cancer molecular subtypes', *Br J Cancer*, 118(8), pp. 1107-1114.

Hanahan, D. and Weinberg, R.A. (2011) 'Hallmarks of cancer: the next generation', in *Cell*. United States: 2011 Elsevier Inc, pp. 646-74.

Harewood, L., Robinson, H., Harris, R., Al-Obaidi, M.J., Jalali, G.R., Martineau, M., Moorman, A.V., Sumption, N., Richards, S., Mitchell, C. and Harrison, C.J. (2003) 'Amplification of AML1 on a duplicated chromosome 21 in acute lymphoblastic leukemia: a study of 20 cases', *Leukemia*, 17(3), pp. 547-53.

Harris, A.W., Pinkert, C.A., Crawford, M., Langdon, W.Y., Brinster, R.L. and Adams, J.M. (1988) 'The E mu-myc transgenic mouse. A model for high-incidence spontaneous lymphoma and leukemia of early B cells', *J Exp Med*, 167(2), pp. 353-71.

Hatzi, K., Jiang, Y., Huang, C., Garrett-Bakelman, F., Gearhart, M.D., Giannopoulou, E.G., Zumbo, P., Kirouac, K., Bhaskara, S., Polo, J.M., Kormaksson, M., MacKerell, A.D., Jr., Xue, F., Mason, C.E., Hiebert, S.W., Prive, G.G., Cerchietti, L., Bardwell, V.J., Elemento, O. and Melnick, A. (2013) 'A hybrid mechanism of action for BCL6 in B cells defined by formation of functionally distinct complexes at enhancers and promoters', *Cell Rep*, 4(3), pp. 578-88.

Havelange, V., Ameye, G., Theate, I., Callet-Bauchu, E., Lippert, E., Luquet, I., Raphael, M., Vikkula, M. and Poirel, H.A. (2016a) 'The peculiar 11q-gain/loss aberration reported in a subset of MYC-negative high-grade B-cell lymphomas can also occur in a MYC-rearranged lymphoma', *Cancer Genet*, 209(3), pp. 117-8.

Havelange, V., Pepermans, X., Ameye, G., Theate, I., Callet-Bauchu, E., Barin, C., Penther, D., Lippert, E., Michaux, L., Mugneret, F., Dastugue, N., Raphael, M., Vikkula, M. and Poirel, H.A. (2016b) 'Genetic differences between paediatric and adult Burkitt lymphomas', *Br J Haematol*.

He, L., Thomson, J.M., Hemann, M.T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S.W., Hannon, G.J. and Hammond, S.M. (2005) 'A microRNA polycistron as a potential human oncogene', *Nature*, 435(7043), pp. 828-33.

Healy, J.A., Nugent, A., Rempel, R.E., Moffitt, A.B., Davis, N.S., Jiang, X., Shingleton, J.R., Zhang, J., Love, C., Datta, J., McKinney, M.E., Tzeng, T.J., Wettschureck, N., Offermanns, S.,

Walzer, K.A., Chi, J.T., Rasheed, S.A., Casey, P.J., Lossos, I.S. and Dave, S.S. (2016) 'GNA13 loss in germinal center B cells leads to impaired apoptosis and promotes lymphoma in vivo', *Blood*, 127(22), pp. 2723-31.

Hesseling, P.B., Broadhead, R., Molyneux, E., Borgstein, E., Schneider, J.W., Louw, M., Mansvelt, E.P. and Wessels, G. (2003) 'Malawi pilot study of Burkitt lymphoma treatment', *Med Pediatr Oncol*, 41(6), pp. 532-40.

Hesseling, P.B., Molyneux, E., Tchintseme, F., Welbeck, J., McCormick, P., Pritchard-Jones, K. and Wagner, H.P. (2008) 'Treating Burkitt's lymphoma in Malawi, Cameroon, and Ghana', in *Lancet Oncol*. England, pp. 512-3.

HMRN - Incidence (2019). Available at: https://www.hmrn.org/statistics/incidence (Accessed: 06/12).

Hollstein, M., Sidransky, D., Vogelstein, B. and Harris, C.C. (1991) 'p53 mutations in human cancers', *Science*, 253(5015), pp. 49-53.

Hummel, M., Bentink, S., Berger, H., Klapper, W., Wessendorf, S., Barth, T.F., Bernd, H.W., Cogliatti, S.B., Dierlamm, J., Feller, A.C., Hansmann, M.L., Haralambieva, E., Harder, L., Hasenclever, D., Kuhn, M., Lenze, D., Lichter, P., Martin-Subero, J.I., Moller, P., Muller-Hermelink, H.K., Ott, G., Parwaresch, R.M., Pott, C., Rosenwald, A., Rosolowski, M., Schwaenen, C., Sturzenhofecker, B., Szczepanowski, M., Trautmann, H., Wacker, H.H., Spang, R., Loeffler, M., Trumper, L., Stein, H. and Siebert, R. (2006) 'A biologic definition of Burkitt's lymphoma from transcriptional and genomic profiling', *N Engl J Med*, 354(23), pp. 2419-30.

Iqbal, J., Greiner, T.C., Patel, K., Dave, B.J., Smith, L., Ji, J., Wright, G., Sanger, W.G., Pickering, D.L., Jain, S., Horsman, D.E., Shen, Y., Fu, K., Weisenburger, D.D., Hans, C.P., Campo, E., Gascoyne, R.D., Rosenwald, A., Jaffe, E.S., Delabie, J., Rimsza, L., Ott, G., Müller-Hermelink, H.K., Connors, J.M., Vose, J.M., McKeithan, T., Staudt, L.M. and Chan, W.C. (2007) 'Distinctive patterns of BCL6 molecular alterations and their functional consequences in different subgroups of diffuse large B-cell lymphoma', *Leukemia*, 21(11), pp. 2332-2343.

Iqbal, J., Sanger, W.G., Horsman, D.E., Rosenwald, A., Pickering, D.L., Dave, B., Dave, S., Xiao, L., Cao, K., Zhu, Q., Sherman, S., Hans, C.P., Weisenburger, D.D., Greiner, T.C., Gascoyne, R.D., Ott, G., Muller-Hermelink, H.K., Delabie, J., Braziel, R.M., Jaffe, E.S., Campo, E., Lynch,

J.C., Connors, J.M., Vose, J.M., Armitage, J.O., Grogan, T.M., Staudt, L.M. and Chan, W.C. (2004) 'BCL2 translocation defines a unique tumor subset within the germinal center B-cell-like diffuse large B-cell lymphoma', *Am J Pathol*, 165(1), pp. 159-66.

Jardin, F., Jais, J.P., Molina, T.J., Parmentier, F., Picquenot, J.M., Ruminy, P., Tilly, H., Bastard, C., Salles, G.A., Feugier, P., Thieblemont, C., Gisselbrecht, C., de Reynies, A., Coiffier, B., Haioun, C. and Leroy, K. (2010) 'Diffuse large B-cell lymphomas with CDKN2A deletion have a distinct gene expression signature and a poor prognosis under R-CHOP treatment: a GELA study', *Blood*, 116(7), pp. 1092-104.

Jiang, Y., Redmond, D., Nie, K., Eng, K.W., Clozel, T., Martin, P., Tan, L.H., Melnick, A.M., Tam, W. and Elemento, O. (2014) 'Deep sequencing reveals clonal evolution patterns and mutation events associated with relapse in B-cell lymphomas', *Genome Biol*, 15(8).

Juskevicius, D., Dirnhofer, S. and Tzankov, A. (2017) 'Genetic background and evolution of relapses in aggressive B-cell lymphomas', *Haematologica*, 102(7), pp. 1139-1149.

Juskevicius, D., Lorber, T., Gsponer, J., Perrina, V., Ruiz, C., Stenner-Liewen, F., Dirnhofer, S. and Tzankov, A. (2016) 'Distinct genetic evolution patterns of relapsing diffuse large B-cell lymphoma revealed by genome-wide copy number aberration and targeted sequencing analysis', *Leukemia*, 30(12), pp. 2385-2395.

Kabrani, E., Chu, V.T., Tasouri, E., Sommermann, T., Bassler, K., Ulas, T., Zenz, T., Bullinger, L., Schultze, J.L., Rajewsky, K. and Sander, S. (2018) 'Nuclear FOXO1 promotes lymphomagenesis in germinal center B cells', *Blood*, 132(25), pp. 2670-2683.

Kamranvar, S.A., Gruhne, B., Szeles, A. and Masucci, M.G. (2007) 'Epstein-Barr virus promotes genomic instability in Burkitt's lymphoma', *Oncogene*, 26(35), pp. 5115-23.

Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. and Tanabe, M. (2019) 'New approach for understanding genome variations in KEGG', *Nucleic Acids Res*, 47(D1), pp. D590-d595.

Kato, L., Begum, N.A., Burroughs, A.M., Doi, T., Kawai, J., Daub, C.O., Kawaguchi, T., Matsuda, F., Hayashizaki, Y. and Honjo, T. (2012) 'Nonimmunoglobulin target loci of activation-induced cytidine deaminase (AID) share unique features with immunoglobulin genes', *Proc Natl Acad Sci U S A*, 109(7), pp. 2479-84.

Kato, M., Sanada, M., Kato, I., Sato, Y., Takita, J., Takeuchi, K., Niwa, A., Chen, Y., Nakazaki, K., Nomoto, J., Asakura, Y., Muto, S., Tamura, A., Iio, M., Akatsuka, Y., Hayashi, Y., Mori, H., Igarashi, T., Kurokawa, M., Chiba, S., Mori, S., Ishikawa, Y., Okamoto, K., Tobinai, K., Nakagama, H., Nakahata, T., Yoshino, T., Kobayashi, Y. and Ogawa, S. (2009) 'Frequent inactivation of A20 in B-cell lymphomas', *Nature*, 459(7247), pp. 712-6.

Kelleher, F.C., Solomon, B. and McArthur, G.A. (2012) 'Molecular therapeutic advances in personalized therapy of melanoma and non-small cell lung cancer', *J Pers Med*, 2(2), pp. 35-49.

Kim, H., Park, E.S., Lee, S.H., Koo, H.H., Kim, H.S., Lyu, C.J., Jun, S.E., Lim, Y.T., Baek, H.J., Kook, H., Lee, J.W., Kang, H.J., Park, K.D., Shin, H.Y. and Ahn, H.S. (2014) 'Clinical Outcome of Relapsed or Refractory Burkitt Lymphoma and Mature B-Cell Lymphoblastic Leukemia in Children and Adolescents', *Cancer Res Treat*, 46(4), pp. 358-65.

Klapper, W., Kreuz, M., Kohler, C.W., Burkhardt, B., Szczepanowski, M., Salaverria, I., Hummel, M., Loeffler, M., Pellissery, S., Woessmann, W., Schwanen, C., Trumper, L., Wessendorf, S., Spang, R., Hasenclever, D. and Siebert, R. (2012) 'Patient age at diagnosis is associated with the molecular characteristics of diffuse large B-cell lymphoma', *Blood*, 119(8), pp. 1882-7.

Kontgen, F., Grumont, R.J., Strasser, A., Metcalf, D., Li, R., Tarlinton, D. and Gerondakis, S. (1995) 'Mice lacking the c-rel proto-oncogene exhibit defects in lymphocyte proliferation, humoral immunity, and interleukin-2 expression', *Genes Dev*, 9(16), pp. 1965-77.

Kresse, S.H., Ohnstad, H.O., Paulsen, E.B., Bjerkehagen, B., Szuhai, K., Serra, M., Schaefer, K.L., Myklebost, O. and Meza-Zepeda, L.A. (2009) 'LSAMP, a novel candidate tumor suppressor gene in human osteosarcomas, identified by array comparative genomic hybridization', *Genes Chromosomes Cancer*, 48(8), pp. 679-93.

Kumar, M., Lu, Z., Takwi, A.A.L., Chen, W., Callander, N.S., Ramos, K.S., Young, K.H. and Li, Y. (2011) 'Negative Regulation of the Tumor Suppressor p53 Gene by MicroRNAs', *Oncogene*, 30(7), pp. 843-53.

Kumar, R., DiMenna, L.J., Chaudhuri, J. and Evans, T. (2014) 'Biological function of activation-induced cytidine deaminase (AID)', *Biomed J*, 37(5), pp. 269-83.

Kwiecinska, A., Ichimura, K., Berglund, M., Dinets, A., Sulaiman, L., Collins, V.P., Larsson, C., Porwit, A. and Lagercrantz, S.B. (2014) 'Amplification of 2p as a genomic marker for transformation in lymphoma', *Genes Chromosomes Cancer*, 53(9), pp. 750-68.

Lam, L.T., Davis, R.E., Pierce, J., Hepperle, M., Xu, Y., Hottelet, M., Nong, Y., Wen, D., Adams, J., Dang, L. and Staudt, L.M. (2005) 'Small molecule inhibitors of IkappaB kinase are selectively toxic for subgroups of diffuse large B-cell lymphoma defined by gene expression profiling', *Clin Cancer Res*, 11(1), pp. 28-40.

Lee, J.W., Soung, Y.H., Kim, S.Y., Nam, S.W., Park, W.S., Lee, J.Y., Yoo, N.J. and Lee, S.H. (2006) 'JAK2 V617F mutation is uncommon in non-Hodgkin lymphomas', *Leuk Lymphoma*, 47(2), pp. 313-4.

Lee, S., Luo, W., Shah, T., Yin, C., O'Connell, T., Chung, T.H., Perkins, S.L., Miles, R.R., Ayello, J., Morris, E., Harrison, L., van de Ven, C. and Cairo, M.S. (2017) 'The effects of DLEU1 gene expression in Burkitt lymphoma (BL): potential mechanism of chemoimmunotherapy resistance in BL', *Oncotarget*, 8(17), pp. 27839-53.

Lenz, G., Wright, G.W., Emre, N.C., Kohlhammer, H., Dave, S.S., Davis, R.E., Carty, S., Lam, L.T., Shaffer, A.L., Xiao, W., Powell, J., Rosenwald, A., Ott, G., Muller-Hermelink, H.K., Gascoyne, R.D., Connors, J.M., Campo, E., Jaffe, E.S., Delabie, J., Smeland, E.B., Rimsza, L.M., Fisher, R.I., Weisenburger, D.D., Chan, W.C. and Staudt, L.M. (2008) 'Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways', *Proc Natl Acad Sci U S A*, 105(36), pp. 13520-5.

Lepretre, F., Villenet, C., Quief, S., Nibourel, O., Jacquemin, C., Troussard, X., Jardin, F., Gibson, F., Kerckaert, J.P., Roumier, C. and Figeac, M. (2010) 'Waved aCGH: to smooth or not to smooth', *Nucleic Acids Res*, 38(7), p. e94.

Leroy, B., Anderson, M. and Soussi, T. (2014) 'TP53 mutations in human cancer: database reassessment and prospects for the next decade', *Hum Mutat*, 35(6), pp. 672-88.

Leucci, E., Cocco, M., Onnis, A., De Falco, G., van Cleef, P., Bellan, C., van Rijk, A., Nyagol, J., Byakika, B., Lazzi, S., Tosi, P., van Krieken, H. and Leoncini, L. (2008) 'MYC translocation-negative classical Burkitt lymphoma cases: an alternative pathogenetic mechanism involving miRNA deregulation', *J Pathol*, 216(4), pp. 440-50.

Leventaki, V., Rodic, V., Tripp, S.R., Bayerl, M.G., Perkins, S.L., Barnette, P., Schiffman, J.D. and Miles, R.R. (2012) 'TP53 pathway analysis in paediatric Burkitt lymphoma reveals increased MDM4 expression as the only TP53 pathway abnormality detected in a subset of cases', *Br J Haematol*, 158(6), pp. 763-71.

Li, F.P., Fraumeni, J.F., Jr., Mulvihill, J.J., Blattner, W.A., Dreyfus, M.G., Tucker, M.A. and Miller, R.W. (1988) 'A cancer family syndrome in twenty-four kindreds', *Cancer Res*, 48(18), pp. 5358-62.

Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows-Wheeler transform', *Bioinformatics*, 25(14), pp. 1754-60.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25(16), pp. 2078-9.

Li, V.D., Li, K.H. and Li, J.T. (2019) 'TP53 mutations as potential prognostic markers for specific cancers: analysis of data from The Cancer Genome Atlas and the International Agency for Research on Cancer TP53 Database', *J Cancer Res Clin Oncol*, 145(3), pp. 625-636.

Limat, S., Demesmay, K., Voillat, L., Bernard, Y., Deconinck, E., Brion, A., Sabbah, A., Woronoff-Lemsi, M.C. and Cahn, J.Y. (2003) 'Early cardiotoxicity of the CHOP regimen in aggressive non-Hodgkin's lymphoma', *Ann Oncol*, 14(2), pp. 277-81.

Lones, M.A., Sanger, W.G., Le Beau, M.M., Heerema, N.A., Sposto, R., Perkins, S.L., Buckley, J., Kadin, M.E., Kjeldsberg, C.R., Meadows, A., Siegel, S., Finlay, J., Bergeron, S. and Cairo, M.S. (2004) 'Chromosome abnormalities may correlate with prognosis in Burkitt/Burkitt-like lymphomas of children and adolescents: a report from Children's Cancer Group Study CCG-E08', *J Pediatr Hematol Oncol*, 26(3), pp. 169-78.

Lopez, C., Kleinheinz, K., Aukema, S.M., Rohde, M., Bernhart, S.H., Hubschmann, D., Wagener, R., Toprak, U.H., Raimondi, F., Kreuz, M., Waszak, S.M., Huang, Z., Sieverling, L., Paramasivam, N., Seufert, J., Sungalee, S., Russell, R.B., Bausinger, J., Kretzmer, H., Ammerpohl, O., Bergmann, A.K., Binder, H., Borkhardt, A., Brors, B., Claviez, A., Doose, G., Feuerbach, L., Haake, A., Hansmann, M.L., Hoell, J., Hummel, M., Korbel, J.O., Lawerenz, C., Lenze, D., Radlwimmer, B., Richter, J., Rosenstiel, P., Rosenwald, A., Schilhabel, M.B., Stein,

H., Stilgenbauer, S., Stadler, P.F., Szczepanowski, M., Weniger, M.A., Zapatka, M., Eils, R., Lichter, P., Loeffler, M., Moller, P., Trumper, L., Klapper, W., Hoffmann, S., Kuppers, R., Burkhardt, B., Schlesner, M. and Siebert, R. (2019) 'Genomic and transcriptomic changes complement each other in the pathogenesis of sporadic Burkitt lymphoma', *Nat Commun*, 10(1), p. 1459.

Love, C., Sun, Z., Jima, D., Li, G., Zhang, J., Miles, R., Richards, K.L., Dunphy, C.H., Choi, W.W., Srivastava, G., Lugar, P.L., Rizzieri, D.A., Lagoo, A.S., Bernal-Mizrachi, L., Mann, K.P., Flowers, C.R., Naresh, K.N., Evens, A.M., Chadburn, A., Gordon, L.I., Czader, M.B., Gill, J.I., Hsi, E.D., Greenough, A., Moffitt, A.B., McKinney, M., Banerjee, A., Grubor, V., Levy, S., Dunson, D.B. and Dave, S.S. (2012) 'The genetic landscape of mutations in Burkitt lymphoma', *Nat Genet*, 44(12), pp. 1321-5.

MacDonald, J.R., Ziman, R., Yuen, R.K., Feuk, L. and Scherer, S.W. (2014) 'The Database of Genomic Variants: a curated collection of structural variation in the human genome', *Nucleic Acids Res*, 42(Database issue), pp. D986-92.

Magrath, I., Adde, M., Shad, A., Venzon, D., Seibel, N., Gootenberg, J., Neely, J., Arndt, C., Nieder, M., Jaffe, E., Wittes, R.A. and Horak, I.D. (1996) 'Adults and children with small non-cleaved-cell lymphoma have a similar excellent outcome when treated with the same chemotherapy regimen', *J Clin Oncol*, 14(3), pp. 925-34.

Malcikova, J., Stano-Kozubik, K., Tichy, B., Kantorova, B., Pavlova, S., Tom, N., Radova, L., Smardova, J., Pardy, F., Doubek, M., Brychtova, Y., Mraz, M., Plevova, K., Diviskova, E., Oltova, A., Mayer, J., Pospisilova, S. and Trbusek, M. (2015) 'Detailed analysis of therapy-driven clonal evolution of TP53 mutations in chronic lymphocytic leukemia', *Leukemia*, 29(4), pp. 877-85.

Malcikova, J., Tausch, E., Rossi, D., Sutton, L.A., Soussi, T., Zenz, T., Kater, A.P., Niemann, C.U., Gonzalez, D., Davi, F., Gonzalez Diaz, M., Moreno, C., Gaidano, G., Stamatopoulos, K., Rosenquist, R., Stilgenbauer, S., Ghia, P. and Pospisilova, S. (2018) 'ERIC recommendations for TP53 mutation analysis in chronic lymphocytic leukemia-update on methodological approaches and results interpretation', *Leukemia*, 32(5), pp. 1070-1080.

Malkin, D., Li, F.P., Strong, L.C., Fraumeni, J.F., Jr., Nelson, C.E., Kim, D.H., Kassel, J., Gryka, M.A., Bischoff, F.Z., Tainsky, M.A. and et al. (1990) 'Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms', *Science*, 250(4985), pp. 1233-8.

Marine, J.C., Francoz, S., Maetens, M., Wahl, G., Toledo, F. and Lozano, G. (2006) 'Keeping p53 in check: essential and synergistic functions of Mdm2 and Mdm4', *Cell Death Differ*, 13(6), pp. 927-34.

Mass, R.D., Press, M.F., Anderson, S., Cobleigh, M.A., Vogel, C.L., Dybdal, N., Leiberman, G. and Slamon, D.J. (2005) 'Evaluation of clinical outcomes according to HER2 detection by fluorescence in situ hybridization in women with metastatic breast cancer treated with trastuzumab', *Clin Breast Cancer*, 6(3), pp. 240-6.

McGoldrick, S.M., Mutyaba, I., Adams, S.V., Larsen, A., Krantz, E.M., Namirembe, C., Mooka, P., Nabakooza, S., Ndagire, M., Mubiru, K., Nabwana, M., Nankinga, R., Gerdts, S., Gordon-Maclean, C., Geriga, F., Omoding, A., Sessle, E., Kambugu, J., Uldrick, T.S., Orem, J. and Casper, C. (2019) 'Survival of children with endemic Burkitt lymphoma in a prospective clinical care project in Uganda', *Pediatr Blood Cancer*, 66(9), p. e27813.

Mehlich, D., Garbicz, F. and Wlodarski, P.K. (2018) 'The emerging roles of the polycistronic miR-106b approximately 25 cluster in cancer - A comprehensive review', *Biomed Pharmacother*, 107, pp. 1183-1195.

Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R. and Getz, G. (2011) 'GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers', in *Genome Biol*. p. R41.

Messori, A., Vaiani, M., Trippoli, S., Rigacci, L., Jerkeman, M. and Longo, G. (2001) 'Survival in patients with intermediate or high grade non-Hodgkin's lymphoma: meta-analysis of randomized studies comparing third generation regimens with CHOP', in *Br J Cancer*. pp. 303-7.

Miles, R.R., Arnold, S. and Cairo, M.S. (2012) 'Risk factors and treatment of childhood and adolescent Burkitt lymphoma/leukaemia', *Br J Haematol*, 156(6), pp. 730-43.

Miles, R.R., Raphael, M., McCarthy, K., Wotherspoon, A., Lones, M.A., Terrier-Lacombe, M.J., Patte, C., Gerrard, M., Auperin, A., Sposto, R., Davenport, V., Cairo, M.S. and Perkins, S.L.

(2008) 'Pediatric Diffuse Large B-cell Lymphoma Demonstrates A High Proliferation Index, Frequent c-Myc Protein Expression, and A High Incidence of Germinal Center Subtype: Report of the French-American-British (FAB) International Study Group', *Pediatr Blood Cancer*, 51(3), pp. 369-74.

Miller, T.P., Grogan, T.M., Dahlberg, S., Spier, C.M., Braziel, R.M., Banks, P.M., Foucar, K., Kjeldsberg, C.R., Levy, N., Nathwani, B.N. and et al. (1994) 'Prognostic significance of the Ki-67-associated proliferative antigen in aggressive non-Hodgkin's lymphomas: a prospective Southwest Oncology Group trial', *Blood*, 83(6), pp. 1460-6.

Minard-Colin, V., Auperin, A., Pillon, M., Burke, A., Anderson, J.R., Barkauskas, D.A., Wheatley, K., Delgado, R., Alexander, S., Uyttebroeck, A., Bollard, C., Zsiros, J., Csoka, M., Goma, G., Tulard, A., Patte, C. and Gross, T.G. (2016) 'Results of the randomized Intergroup trial Inter-B-NHL Ritux 2010 for children and adolescents with high-risk B-cell non-Hodgkin lymphoma (B-NHL) and mature acute leukemia (B-AL): Evaluation of rituximab (R) efficacy in addition to standard LMB chemotherapy (CT) regimen', *Journal of Clinical Oncology*, 34(15_suppl), pp. 10507-10507.

Molyneux, E., Schwalbe, E., Chagaluka, G., Banda, K., Israels, T., Depani, S., Mittermayer-Vassallo, K., Windebank, K., Mvula, J., Njiram'madzi, J., O'Brien, S., Carey, P. and Bailey, S. (2017) 'The use of anthracyclines in the treatment of endemic Burkitt lymphoma', *Br J Haematol*, 177(6), pp. 984-990.

Molyneux, E.M., Rochford, R., Griffin, B., Newton, R., Jackson, G., Menon, G., Harrison, C.J., Israels, T. and Bailey, S. (2012) 'Burkitt's lymphoma', in *Lancet*. England: 2012 Elsevier Ltd, pp. 1234-44.

Moorman, A.V. (2016) 'New and emerging prognostic and predictive genetic biomarkers in B-cell precursor acute lymphoblastic leukemia', *Haematologica*, 101(4), pp. 407-16.

Moormann, A.M. and Bailey, J.A. (2016) 'Malaria – how this parasitic infection aids and abets EBV-associated Burkitt lymphomagenesis', *Curr Opin Virol*, 20, pp. 78-84.

Nelson, M., Perkins, S.L., Dave, B.J., Coccia, P.F., Bridge, J.A., Lyden, E.R., Heerema, N.A., Lones, M.A., Harrison, L., Cairo, M.S. and Sanger, W.G. (2010) 'An increased frequency of 13q deletions detected by fluorescence in situ hybridization and its impact on survival in children

and adolescents with Burkitt lymphoma: results from the Children's Oncology Group study CCG-5961', *Br J Haematol*, 148(4), pp. 600-10.

Ngo, V.N., Davis, R.E., Lamy, L., Yu, X., Zhao, H., Lenz, G., Lam, L.T., Dave, S., Yang, L., Powell, J. and Staudt, L.M. (2006) 'A loss-of-function RNA interference screen for molecular targets in cancer', *Nature*, 441(7089), pp. 106-10.

Ntougkos, E., Rush, R., Scott, D., Frankenberg, T., Gabra, H., Smyth, J.F. and Sellar, G.C. (2005) 'The IgLON family in epithelial ovarian cancer: expression profiles and clinicopathologic correlates', *Clin Cancer Res*, 11(16), pp. 5764-8.

O'Callaghan-Gordo, C., Casabonne, D., Carrilho, C., Ferro, J., Lorenzoni, C., Zaqueu, C., Nhabomba, A., Aguilar, R., Bassat, Q., de Sanjosé, S., Dobaño, C. and Kogevinas, M. (2016) 'Incidence of Endemic Burkitt Lymphoma in Three Regions of Mozambique', *Am J Trop Med Hyg*, 95(6), pp. 1459-62.

O'Leary, SM, M., A, V. and ME, P. (2004) 'Mdm2 regulates p53 independently of p19(ARF) in homeostatic tissues', *Molecular and cellular biology*, 24(1).

Olivier, M., Goldgar, D.E., Sodha, N., Ohgaki, H., Kleihues, P., Hainaut, P. and Eeles, R.A. (2003) 'Li-Fraumeni and related syndromes: correlation between tumor type, family structure, and TP53 genotype', *Cancer Res*, 63(20), pp. 6643-50.

Olivier, M., Hollstein, M. and Hainaut, P. (2010) 'TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use', in *Cold Spring Harb Perspect Biol*.

Onciu, M., Schlette, E., Zhou, Y., Raimondi, S.C., Giles, F.J., Kantarjian, H.M., Medeiros, L.J., Ribeiro, R.C., Pui, C.H. and Sandlund, J.T. (2006) 'Secondary chromosomal abnormalities predict outcome in pediatric and adult high-stage Burkitt lymphoma', *Cancer*, 107(5), pp. 1084-92.

Opitz, L., Tan, G., Aquino, C., Schlapbach, R. and Rehrauer, H. (2019) 'Long Fragments Achieve Lower Base Quality in Illumina Paired-End Sequencing', *J Biomol Tech*, 30(Suppl), p. S26.

Oschlies, I., Klapper, W., Zimmermann, M., Krams, M., Wacker, H.-H., Burkhardt, B., Harder, L., Siebert, R., Reiter, A. and Parwaresch, R. (2006) 'Diffuse large B-cell lymphoma in pediatric

patients belongs predominantly to the germinal-center type B-cell lymphomas: a clinicopathologic analysis of cases included in the German BFM (Berlin-Frankfurt-Münster) Multicenter Trial', 107(10), pp. 4047-4052.

O'Hayre, M., Inoue, A., Kufareva, I., Wang, Z., Mikelis, C.M., Drummond, R.A., Avino, S., Finkel, K., Kalim, K., DiPasquale, G., Guo, F., Aoki, J., Zheng, Y., Lionakis, M.S., Molinolo, A.A. and Gutkind, J.S. (2016) 'Inactivating Mutations in GNA13 and RHOA in Burkitt's Lymphoma and Diffuse Large B cell Lymphoma: A Tumor Suppressor Function for the Gα13/RhoA Axis in B Cells', *Oncogene*, 35(29), pp. 3771-80.

Panea, R.I., Love, C.L., Shingleton, J.R., Reddy, A., Bailey, J.A., Moormann, A.M., Otieno, J.A., Ong'echa, J.M., Oduor, C.I., Schroeder, K.M.S., Masalu, N., Chao, N.J., Agajanian, M., Major, M.B., Fedoriw, Y., Richards, K.L., Rymkiewicz, G., Miles, R.R., Alobeid, B., Bhagat, G., Flowers, C.R., Ondrejka, S.L., Hsi, E.D., Choi, W.W.L., Au-Yeung, R.K.H., Hartmann, W., Lenz, G., Meyerson, H., Lin, Y.Y., Zhuang, Y., Luftig, M.A., Waldrop, A., Dave, T., Thakkar, D., Sahay, H., Li, G., Palus, B.C., Seshadri, V., Kim, S.Y., Gascoyne, R.D., Levy, S., Mukhopadyay, M., Dunson, D.B. and Dave, S.S. (2019) 'The whole-genome landscape of Burkitt lymphoma subtypes', *Blood*, 134(19), pp. 1598-1607.

Pasqualucci, L., Compagno, M., Houldsworth, J., Monti, S., Grunn, A., Nandula, S.V., Aster, J.C., Murty, V.V., Shipp, M.A. and Dalla-Favera, R. (2006) 'Inactivation of the PRDM1/BLIMP1 gene in diffuse large B cell lymphoma', in *J Exp Med*. pp. 311-7.

Patte, C., Auperin, A., Gerrard, M., Michon, J., Pinkerton, R., Sposto, R., Weston, C., Raphael, M., Perkins, S.L., McCarthy, K. and Cairo, M.S. (2007) 'Results of the randomized international FAB/LMB96 trial for intermediate risk B-cell non-Hodgkin lymphoma in children and adolescents: it is possible to reduce treatment for the early responding patients', *Blood*, 109(7), pp. 2773-80.

Pearson, A.D.J., Scobie, N., Norga, K., Ligas, F., Chiodin, D., Burke, A., Minard-Colin, V., Adamson, P., Marshall, L.V., Balakumaran, A., Benettaib, B., Bhargava, P., Bollard, C.M., Bolotin, E., Bomken, S., Buechner, J., Burkhardt, B., Caron, H., Copland, C., Demolis, P., Egorov, A., Farhan, M., Zugmaier, G., Gross, T., Horton-Taylor, D., Klapper, W., Lesa, G., Marcus, R., Miles, R.R., Nottage, K., Pacaud, L., Ricafort, R., Schrappe, M., Sterba, J., Vezan, R., Weiner, S., Kim, S.Y., Reaman, G. and Vassal, G. (2019) 'ACCELERATE and European

Medicine Agency Paediatric Strategy Forum for medicinal product development for mature B-cell malignancies in children', *Eur J Cancer*, 110, pp. 74-85.

Pelicci, P.G., Knowles, D.M., Magrath, I. and Dalla-Favera, R. (1986) 'Chromosomal breakpoints and structural alterations of the c-myc locus differ in endemic and sporadic forms of Burkitt lymphoma', *Proc Natl Acad Sci U S A*, 83(9), pp. 2984-8.

Perkins, A.S. and Friedberg, J.W. (2008) 'Burkitt lymphoma in adults', *Hematology Am Soc Hematol Educ Program*, pp. 341-8.

Petitjean, A., Mathe, E., Kato, S., Ishioka, C., Tavtigian, S.V., Hainaut, P. and Olivier, M. (2007) 'Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database', *Hum Mutat*, 28(6), pp. 622-9.

Pfeifer, M., Grau, M., Lenze, D., Wenzel, S.S., Wolf, A., Wollert-Wulf, B., Dietze, K., Nogai, H., Storek, B., Madle, H., Dorken, B., Janz, M., Dirnhofer, S., Lenz, P., Hummel, M., Tzankov, A. and Lenz, G. (2013) 'PTEN loss defines a PI3K/AKT pathway-dependent germinal center subtype of diffuse large B-cell lymphoma', *Proc Natl Acad Sci U S A*, 110(30), pp. 12420-5.

Pfreundschuh, M., Trumper, L., Osterborg, A., Pettengell, R., Trneny, M., Imrie, K., Ma, D., Gill, D., Walewski, J., Zinzani, P.L., Stahel, R., Kvaloy, S., Shpilberg, O., Jaeger, U., Hansen, M., Lehtinen, T., Lopez-Guillermo, A., Corrado, C., Scheliga, A., Milpied, N., Mendila, M., Rashford, M., Kuhnt, E. and Loeffler, M. (2006) 'CHOP-like chemotherapy plus rituximab versus CHOP-like chemotherapy alone in young patients with good-prognosis diffuse large-B-cell lymphoma: a randomised controlled trial by the MabThera International Trial (MInT) Group', *Lancet Oncol*, 7(5), pp. 379-91.

Phillips, D.C., Xiao, Y., Lam, L.T., Litvinovich, E., Roberts-Rapp, L., Souers, A.J. and Leverson, J.D. (2015) 'Loss in MCL-1 function sensitizes non-Hodgkin's lymphoma cell lines to the BCL-2-selective inhibitor venetoclax (ABT-199)', *Blood Cancer J*, 5, p. e368.

Poirel, H.A., Cairo, M.S., Heerema, N.A., Swansbury, J., Auperin, A., Launay, E., Sanger, W.G., Talley, P., Perkins, S.L., Raphael, M., McCarthy, K., Sposto, R., Gerrard, M., Bernheim, A. and Patte, C. (2009) 'Specific cytogenetic abnormalities are associated with a significantly inferior

outcome in children and adolescents with mature B-cell non-Hodgkin's lymphoma: results of the FAB/LMB 96 international study', *Leukemia*, 23(2), pp. 323-31.

Quinlan, A.R. and Hall, I.M. (2010) 'BEDTools: a flexible suite of utilities for comparing genomic features', *Bioinformatics*, 26(6), pp. 841-2.

Rand, V., Parker, H., Russell, L.J., Schwab, C., Ensor, H., Irving, J., Jones, L., Masic, D., Minto, L., Morrison, H., Ryan, S., Robinson, H., Sinclair, P., Moorman, A.V., Strefford, J.C. and Harrison, C.J. (2011) 'Genomic characterization implicates iAMP21 as a likely primary genetic event in childhood B-cell precursor acute lymphoblastic leukemia', *Blood*, 117(25), pp. 6848-55.

Rausch, T., Jones, D.T., Zapatka, M., Stutz, A.M., Zichner, T., Weischenfeldt, J., Jager, N., Remke, M., Shih, D., Northcott, P.A., Pfaff, E., Tica, J., Wang, Q., Massimi, L., Witt, H., Bender, S., Pleier, S., Cin, H., Hawkins, C., Beck, C., von Deimling, A., Hans, V., Brors, B., Eils, R., Scheurlen, W., Blake, J., Benes, V., Kulozik, A.E., Witt, O., Martin, D., Zhang, C., Porat, R., Merino, D.M., Wasserman, J., Jabado, N., Fontebasso, A., Bullinger, L., Rucker, F.G., Dohner, K., Dohner, H., Koster, J., Molenaar, J.J., Versteeg, R., Kool, M., Tabori, U., Malkin, D., Korshunov, A., Taylor, M.D., Lichter, P., Pfister, S.M. and Korbel, J.O. (2012) 'Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations', *Cell*, 148(1-2), pp. 59-71.

Ren, B., Yu, G., Tseng, G.C., Cieply, K., Gavel, T., Nelson, J., Michalopoulos, G., Yu, Y.P. and Luo, J.H. (2006) 'MCM7 amplification and overexpression are associated with prostate cancer progression', *Oncogene*, 25(7), pp. 1090-8.

Richter, J., Schlesner, M., Hoffmann, S., Kreuz, M., Leich, E., Burkhardt, B., Rosolowski, M., Ammerpohl, O., Wagener, R., Bernhart, S.H., Lenze, D., Szczepanowski, M., Paulsen, M., Lipinski, S., Russell, R.B., Adam-Klages, S., Apic, G., Claviez, A., Hasenclever, D., Hovestadt, V., Hornig, N., Korbel, J.O., Kube, D., Langenberger, D., Lawerenz, C., Lisfeld, J., Meyer, K., Picelli, S., Pischimarov, J., Radlwimmer, B., Rausch, T., Rohde, M., Schilhabel, M., Scholtysik, R., Spang, R., Trautmann, H., Zenz, T., Borkhardt, A., Drexler, H.G., Moller, P., MacLeod, R.A., Pott, C., Schreiber, S., Trumper, L., Loeffler, M., Stadler, P.F., Lichter, P., Eils, R., Kuppers, R., Hummel, M., Klapper, W., Rosenstiel, P., Rosenwald, A., Brors, B. and Siebert, R. (2012)

'Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing', *Nat Genet*, 44(12), pp. 1316-20.

Rieber, N., Bohnert, R., Ziehm, U. and Jansen, G. (2017) 'Reliability of algorithmic somatic copy number alteration detection from targeted capture data', *Bioinformatics*, 33(18), pp. 2791-2798.

Robinson, H.M., Harrison, C.J., Moorman, A.V., Chudoba, I. and Strefford, J.C. (2007) 'Intrachromosomal amplification of chromosome 21 (iAMP21) may arise from a breakage-fusion-bridge cycle', *Genes Chromosomes Cancer*, 46(4), pp. 318-26.

Robles, A.I., Jen, J. and Harris, C.C. (2016) 'Clinical Outcomes of TP53 Mutations in Cancers', in *Cold Spring Harb Perspect Med*.

Rosenwald, A., Wright, G., Chan, W.C., Connors, J.M., Campo, E., Fisher, R.I., Gascoyne, R.D., Muller-Hermelink, H.K., Smeland, E.B., Giltnane, J.M., Hurt, E.M., Zhao, H., Averett, L., Yang, L., Wilson, W.H., Jaffe, E.S., Simon, R., Klausner, R.D., Powell, J., Duffey, P.L., Longo, D.L., Greiner, T.C., Weisenburger, D.D., Sanger, W.G., Dave, B.J., Lynch, J.C., Vose, J., Armitage, J.O., Montserrat, E., Lopez-Guillermo, A., Grogan, T.M., Miller, T.P., LeBlanc, M., Ott, G., Kvaloy, S., Delabie, J., Holte, H., Krajci, P., Stokke, T. and Staudt, L.M. (2002) 'The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma', *N Engl J Med*, 346(25), pp. 1937-47.

Rossi, D., Khiabanian, H., Spina, V., Ciardullo, C., Bruscaggin, A., Fama, R., Rasi, S., Monti, S., Deambrogi, C., De Paoli, L., Wang, J., Gattei, V., Guarini, A., Foa, R., Rabadan, R. and Gaidano, G. (2014) 'Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia', *Blood*, 123(14), pp. 2139-47.

Rucker, F.G., Schlenk, R.F., Bullinger, L., Kayser, S., Teleanu, V., Kett, H., Habdank, M., Kugler, C.M., Holzmann, K., Gaidzik, V.I., Paschka, P., Held, G., von Lilienfeld-Toal, M., Lubbert, M., Frohling, S., Zenz, T., Krauter, J., Schlegelberger, B., Ganser, A., Lichter, P., Dohner, K. and Dohner, H. (2012) 'TP53 alterations in acute myeloid leukemia with complex karyotype correlate with specific copy number alterations, monosomal karyotype, and dismal outcome', *Blood*, 119(9), pp. 2114-21.

Saha, A., Murakami, M., Kumar, P., Bajaj, B., Sims, K. and Robertson, E.S. (2009) 'Epstein-Barr virus nuclear antigen 3C augments Mdm2-mediated p53 ubiquitination and degradation by deubiquitinating Mdm2', *J Virol*, 83(9), pp. 4652-69.

Salaverria, I., Martin-Guerrero, I., Wagener, R., Kreuz, M., Kohler, C.W., Richter, J., Pienkowska-Grela, B., Adam, P., Burkhardt, B., Claviez, A., Damm-Welk, C., Drexler, H.G., Hummel, M., Jaffe, E.S., Kuppers, R., Lefebvre, C., Lisfeld, J., Loffler, M., Macleod, R.A., Nagel, I., Oschlies, I., Rosolowski, M., Russell, R.B., Rymkiewicz, G., Schindler, D., Schlesner, M., Scholtysik, R., Schwaenen, C., Spang, R., Szczepanowski, M., Trumper, L., Vater, I., Wessendorf, S., Klapper, W. and Siebert, R. (2014) 'A recurrent 11q aberration pattern characterizes a subset of MYC-negative high-grade B-cell lymphomas resembling Burkitt lymphoma', *Blood*, 123(8), pp. 1187-98.

Salaverria, I., Philipp, C., Oschlies, I., Kohler, C.W., Kreuz, M., Szczepanowski, M., Burkhardt, B., Trautmann, H., Gesk, S., Andrusiewicz, M., Berger, H., Fey, M., Harder, L., Hasenclever, D., Hummel, M., Loeffler, M., Mahn, F., Martin-Guerrero, I., Pellissery, S., Pott, C., Pfreundschuh, M., Reiter, A., Richter, J., Rosolowski, M., Schwaenen, C., Stein, H., Trumper, L., Wessendorf, S., Spang, R., Kuppers, R., Klapper, W. and Siebert, R. (2011) 'Translocations activating IRF4 identify a subtype of germinal center-derived B-cell lymphoma affecting predominantly children and young adults', *Blood*, 118(1), pp. 139-47.

Sande, C.M., Chang, B., Monga, V., Bossler, A.D. and Ma, D. (2018) 'Biallelic TP53 gain of function mutations in rapidly progressing solid tumors', *Cancer Genet*, 222-223, pp. 20-24.

Sander, S., Calado, D.P., Srinivasan, L., Kochert, K., Zhang, B., Rosolowski, M., Rodig, S.J., Holzmann, K., Stilgenbauer, S., Siebert, R., Bullinger, L. and Rajewsky, K. (2012) 'Synergy between PI3K signaling and MYC in Burkitt lymphomagenesis', *Cancer Cell*, 22(2), pp. 167-79.

Sandlund, J.T. and Martin, M.G. (2016) 'Non-Hodgkin lymphoma across the pediatric and adolescent and young adult age spectrum', *Hematology Am Soc Hematol Educ Program*, 2016(1), pp. 589-97.

Savage, K.J., Johnson, N.A., Ben-Neriah, S., Connors, J.M., Sehn, L.H., Farinha, P., Horsman, D.E. and Gascoyne, R.D. (2009) 'MYC gene rearrangements are associated with a poor prognosis in diffuse large B-cell lymphoma patients treated with R-CHOP chemotherapy', *Blood*, 114(17), pp. 3533-7.

Schiffman, J.D., Lorimer, P.D., Rodic, V., Jahromi, M.S., Downie, J.M., Bayerl, M.G., Sanmann, J.N., Althof, P.A., Sanger, W.G., Barnette, P., Perkins, S.L. and Miles, R.R. (2011) 'Genome wide copy number analysis of paediatric Burkitt lymphoma using formalin-fixed tissues reveals a subset with gain of chromosome 13q and corresponding miRNA over expression', *Br J Haematol*, 155(4), pp. 477-86.

Schmitz, R., Young, R.M., Ceribelli, M., Jhavar, S., Xiao, W., Zhang, M., Wright, G., Shaffer, A.L., Hodson, D.J., Buras, E., Liu, X., Powell, J., Yang, Y., Xu, W., Zhao, H., Kohlhammer, H., Rosenwald, A., Kluin, P., Müller-Hermelink, H.K., Ott, G., Gascoyne, R.D., Connors, J.M., Rimsza, L.M., Campo, E., Jaffe, E.S., Delabie, J., Smeland, E.B., Ogwang, M.D., Reynolds, S.J., Fisher, R.I., Braziel, R.M., Tubbs, R.R., Cook, J.R., Weisenburger, D.D., Chan, W.C., Pittaluga, S., Wilson, W., Waldmann, T.A., Rowe, M., Mbulaiteye, S.M., Rickinson, A.B. and Staudt, L.M. (2012) 'Burkitt Lymphoma Pathogenesis and Therapeutic Targets from Structural and Functional Genomics', *Nature*, 490(7418), pp. 116-20.

Scholtysik, R., Kreuz, M., Hummel, M., Rosolowski, M., Szczepanowski, M., Klapper, W., Loeffler, M., Trumper, L., Siebert, R. and Kuppers, R. (2015) 'Characterization of genomic imbalances in diffuse large B-cell lymphoma by detailed SNP-chip analysis', *Int J Cancer*, 136(5), pp. 1033-42.

Scholtysik, R., Kreuz, M., Klapper, W., Burkhardt, B., Feller, A.C., Hummel, M., Loeffler, M., Rosolowski, M., Schwaenen, C., Spang, R., Stein, H., Thorns, C., Trumper, L., Vater, I., Wessendorf, S., Zenz, T., Siebert, R. and Kuppers, R. (2010) 'Detection of genomic aberrations in molecularly defined Burkitt's lymphoma by array-based, high resolution, single nucleotide polymorphism analysis', *Haematologica*, 95(12), pp. 2047-55.

Scholtysik, R., Nagel, I., Kreuz, M., Vater, I., Giefing, M., Schwaenen, C., Wessendorf, S., Trumper, L., Loeffler, M., Siebert, R. and Kuppers, R. (2012) 'Recurrent deletions of the TNFSF7 and TNFSF9 genes in 19p13.3 in diffuse large B-cell and Burkitt lymphomas', *Int J Cancer*, 131(5), pp. E830-5.

Schwalbe, E.C., Lindsey, J.C., Nakjang, S., Crosier, S., Smith, A.J., Hicks, D., Rafiee, G., Hill, R.M., Iliasova, A., Stone, T., Pizer, B., Michalski, A., Joshi, A., Wharton, S.B., Jacques, T.S., Bailey, S., Williamson, D. and Clifford, S.C. (2017) 'Novel molecular subgroups for clinical

classification and outcome prediction in childhood medulloblastoma: a cohort study', *Lancet Oncol*, 18(7), pp. 958-971.

Sehn, L.H. (2012) 'Paramount prognostic factors that guide therapeutic strategies in diffuse large B-cell lymphoma', *Hematology Am Soc Hematol Educ Program*, 2012, pp. 402-9.

Sevilla, D.W., Gong, J.Z., Goodman, B.K., Buckley, P.J., Rosoff, P., Gockerman, J.P. and Lagoo, A.S. (2007) 'Clinicopathologic findings in high-grade B-cell lymphomas with typical Burkitt morphologic features but lacking the MYC translocation', *Am J Clin Pathol*, 128(6), pp. 981-91.

Shaffer, A.L., Lin, K.I., Kuo, T.C., Yu, X., Hurt, E.M., Rosenwald, A., Giltnane, J.M., Yang, L., Zhao, H., Calame, K. and Staudt, L.M. (2002) 'Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program', *Immunity*, 17(1), pp. 51-62.

Shen, W.H., Balajee, A.S., Wang, J., Wu, H., Eng, C., Pandolfi, P.P. and Yin, Y. (2007) 'Essential role for nuclear PTEN in maintaining chromosomal integrity', *Cell*, 128(1), pp. 157-70.

Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001) 'dbSNP: the NCBI database of genetic variation', *Nucleic Acids Res*, 29(1), pp. 308-11.

Sinclair, P.B., Sorour, A., Martineau, M., Harrison, C.J., Mitchell, W.A., O'Neill, E. and Foroni, L. (2004) 'A fluorescence in situ hybridization map of 6q deletions in acute lymphocytic leukemia: identification and analysis of a candidate tumor suppressor gene', *Cancer Res*, 64(12), pp. 4089-98.

Singh, R. and Mukhopadhyay, K. (2011) 'Survival analysis in clinical trials: Basics and must know areas', *Perspect Clin Res*, 2(4), pp. 145-8.

Sobol, H., Benziane, A., Kerangueven, F., Yin, L., Noguchi, T., Pauly, S., Eisinger, F., Longy, M., Romeo, G., Lenoir, G. and Birnbaum, D. (2002) 'Genome-wide search for loss of heterozygosity in Burkitt lymphoma cell lines', *Genes Chromosomes Cancer*, 33(2), pp. 217-24.

Soussi, T. and Wiman, K.G. (2015) 'TP53: an oncogene in disguise', *Cell Death Differ*, 22(8), pp. 1239-49.

Stilgenbauer, S., Eichhorst, B., Schetelig, J., Coutre, S., Seymour, J.F., Munir, T., Puvvada, S.D., Wendtner, C.M., Roberts, A.W., Jurczak, W., Mulligan, S.P., Bottcher, S., Mobasher, M., Zhu, M., Desai, M., Chyla, B., Verdugo, M., Enschede, S.H., Cerri, E., Humerickhouse, R., Gordon, G., Hallek, M. and Wierda, W.G. (2016) 'Venetoclax in relapsed or refractory chronic lymphocytic leukaemia with 17p deletion: a multicentre, open-label, phase 2 study', *Lancet Oncol*, 17(6), pp. 768-778.

Stone, S., Dayananth, P., Jiang, P., Weaver-Feldhaus, J.M., Tavtigian, S.V., Cannon-Albright, L. and Kamb, A. (1995) 'Genomic structure, expression and mutational analysis of the P15 (MTS2) gene', *Oncogene*, 11(5), pp. 987-91.

Su, I.H., Basavaraj, A., Krutchinsky, A.N., Hobert, O., Ullrich, A., Chait, B.T. and Tarakhovsky, A. (2003) 'Ezh2 controls B cell development through histone H3 methylation and Igh rearrangement', *Nat Immunol*, 4(2), pp. 124-31.

Suguro, M., Yoshida, N., Umino, A., Kato, H., Tagawa, H., Nakagawa, M., Fukuhara, N., Karnan, S., Takeuchi, I., Hocking, T.D., Arita, K., Karube, K., Tsuzuki, S., Nakamura, S., Kinoshita, T. and Seto, M. (2014) 'Clonal heterogeneity of lymphoid malignancies correlates with poor prognosis', *Cancer Sci*, 105(7), pp. 897-904.

Swaminathan, S., Huang, C., Geng, H., Chen, Z., Harvey, R., Kang, H., Ng, C., Titz, B., Hurtz, C., Sadiyah, M.F., Nowak, D., Thoennissen, G.B., Rand, V., Graeber, T.G., Koeffler, H.P., Carroll, W.L., Willman, C.L., Hall, A.G., Igarashi, K., Melnick, A. and Muschen, M. (2013) 'BACH2 mediates negative selection and p53-dependent tumor suppression at the pre-B cell receptor checkpoint', *Nat Med*, 19(8), pp. 1014-22.

Swerdlow, S.H., Campo, E., Pileri, S.A., Harris, N.L., Stein, H., Siebert, R., Advani, R., Ghielmini, M., Salles, G.A., Zelenetz, A.D. and Jaffe, E.S. (2016) 'The 2016 revision of the World Health Organization classification of lymphoid neoplasms', *Blood*, 127(20), pp. 2375-90.

Tagawa, H., Karube, K., Tsuzuki, S., Ohshima, K. and Seto, M. (2007) 'Synergistic action of the microRNA-17 polycistron and Myc in aggressive cancer development', *Cancer Sci*, 98(9), pp. 1482-90.

Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S.C., Kok, C.Y., Noble, K., Ponting, L., Ramshaw, C.C., Rye, C.E., Speedy, H.E., Stefancsik, R., Thompson, S.L., Wang, S., Ward, S., Campbell, P.J. and Forbes, S.A. (2019) 'COSMIC: the Catalogue Of Somatic Mutations In Cancer', *Nucleic Acids Res*, 47(D1), pp. D941-d947.

Tian, S., Yan, H., Kalmbach, M. and Slager, S.L. (2016) 'Impact of post-alignment processing in variant discovery from whole exome data', *BMC Bioinformatics*, 17.

Tisato, V., Voltan, R., Gonelli, A., Secchiero, P. and Zauli, G. (2017) 'MDM2/X inhibitors under clinical evaluation: perspectives for the management of hematological malignancies and pediatric cancer', in *J Hematol Oncol*.

Toujani, S., Dessen, P., Ithzar, N., Danglot, G., Richon, C., Vassetzky, Y., Robert, T., Lazar, V., Bosq, J., Da Costa, L., Pérot, C., Ribrag, V., Patte, C., Wiels, J. and Bernheim, A. (2009) 'High Resolution Genome-Wide Analysis of Chromosomal Alterations in Burkitt's Lymphoma', in *PLoS One*.

Trinh, D.L., Scott, D.W., Morin, R.D., Mendez-Lago, M., An, J., Jones, S.J.M., Mungall, A.J., Zhao, Y., Schein, J., Steidl, C., Connors, J.M., Gascoyne, R.D. and Marra, M.A. (2013) 'Analysis of FOXO1 mutations in diffuse large B-cell lymphoma', *Blood*, 121(18), pp. 3666-74.

V, M.-C., A, A., M, P., GAA, B., DA, B., K, W., RF, D., S, A., A, U., CM, B., J, Z., M, C., B, K., AK, C., RR, M., A, W., PC, A., G, V., C, P. and TG, G. (2020) 'Rituximab for High-Risk, Mature B-Cell Non-Hodgkin's Lymphoma in Children', *The New England journal of medicine*, 382(23).

van der Krogt, J.A., Bempt, M.V., Ferreiro, J.F., Mentens, N., Jacobs, K., Pluys, U., Doms, K., Geerdens, E., Uyttebroeck, A., Pierre, P., Michaux, L., Devos, T., Vandenberghe, P., Tousseyn, T., Cools, J. and Wlodarska, I. (2017) 'Anaplastic lymphoma kinase-positive anaplastic large cell lymphoma with the variant RNF213-, ATIC- and TPM3-ALK fusions is characterized by copy number gain of the rearranged ALK gene', *Haematologica*, 102(9), pp. 1605-1616.

van Rijk, A., Mason, D., Jones, M., Cabecadas, J., Crespo, M., Cigudosa, J.C., Garcia, J.F., Leoncini, L., Cocco, M., Hansmann, M.L., Mottok, A., Copie Bergman, C., Baia, M., Anagnostou, D., Pouliou, E., Hamilton Dutoit, S., Hjollund Christiansen, M., Svenstrup Poulsen, T., Hauge Matthiesen, S., van Dongen, J. and van Krieken, J.H. (2008) 'Translocation

detection in lymphoma diagnosis by split-signal FISH: a standardised approach', *J Hematop*, 1(2), pp. 119-26.

Vance, K.W., Carreira, S., Brosch, G. and Goding, C.R. (2005) 'Tbx2 is overexpressed and plays an important role in maintaining proliferation and suppression of senescence in melanomas', *Cancer Res*, 65(6), pp. 2260-8.

Victora, G.D., Schwickert, T.A., Fooksman, D.R., Kamphorst, A.O., Meyer-Hermann, M., Dustin, M.L. and Nussenzweig, M.C. (2010) 'Germinal center dynamics revealed by multiphoton microscopy with a photoactivatable fluorescent reporter', *Cell*, 143(4), pp. 592-605.

Visco, C., Tzankov, A., Xu-Monette, Z.Y., Miranda, R.N., Tai, Y.C., Li, Y., Liu, W., d'Amore, E.S.G., Montes-Moreno, S., Dybkær, K., Chiu, A., Orazi, A., Zu, Y., Bhagat, G., Wang, H.Y., Dunphy, C.H., His, E.D., Zhao, X.F., Choi, W.W., Zhao, X., van Krieken, J.H., Huang, Q., Ai, W., O'Neill, S., Ponzoni, M., Ferreri, A.J., Kahl, B.S., Winter, J.N., Go, R.S., Dirnhofer, S., Piris, M.A., Møller, M.B., Wu, L., Medeiros, L.J. and Young, K.H. (2013) 'Patients with diffuse large B-cell lymphoma of germinal center origin with BCL2 translocations have poor outcome, irrespective of MYC status: a report from an International DLBCL rituximab-CHOP Consortium Program Study', *Haematologica*, 98(2), pp. 255-63.

Vrana, J.A., Bieszczad, C.K., Cleaveland, E.S., Ma, Y., Park, J.P., Mohandas, T.K. and Craig, R.W. (2002) 'An MCL1-overexpressing Burkitt lymphoma subline exhibits enhanced survival on exposure to serum deprivation, topoisomerase inhibitors, or staurosporine but remains sensitive to 1-beta-D-arabinofuranosylcytosine', *Cancer Res*, 62(3), pp. 892-900.

Yamamoto, S. and Iwakuma, T. (2019) 'Regulators of Oncogenic Mutant TP53 Gain of Function', *Cancers (Basel)*, 11(1).

Yip, K.W. and Reed, J.C. (2008) 'Bcl-2 family proteins and cancer', *Oncogene*, 27(50), pp. 6398-6406.

Zenz, T., Habe, S., Denzel, T., Mohr, J., Winkler, D., Buhler, A., Sarno, A., Groner, S., Mertens, D., Busch, R., Hallek, M., Dohner, H. and Stilgenbauer, S. (2009) 'Detailed analysis of p53 pathway defects in fludarabine-refractory chronic lymphocytic leukemia (CLL): dissecting the

contribution of 17p deletion, TP53 mutation, p53-p21 dysfunction, and miR34a in a prospective clinical trial', *Blood*, 114(13), pp. 2589-97.

Zhang, G., Wang, Z., Qian, F., Zhao, C. and Sun, C. (2015) 'Silencing of the ABCC4 gene by RNA interference reverses multidrug resistance in human gastric cancer', *Oncol Rep*, 33(3), pp. 1147-54.

Zhang, L., Zhou, F., Garcia de Vinuesa, A., de Kruijf, E.M., Mesker, W.E., Hui, L., Drabsch, Y., Li, Y., Bauer, A., Rousseau, A., Sheppard, K.A., Mickanin, C., Kuppen, P.J., Lu, C.X. and Ten Dijke, P. (2013) 'TRAF4 promotes TGF-beta receptor signaling and drives breast cancer metastasis', *Mol Cell*, 51(5), pp. 559-72.

Zhang, M., Zhuang, G., Sun, X., Shen, Y., Wang, W., Li, Q. and Di, W. (2017) 'TP53 mutation-mediated genomic instability induces the evolution of chemoresistance and recurrence in epithelial ovarian cancer', *Diagn Pathol*, 12.

Zhou, P., Blain, A.E., Newman, A.M., Zaka, M., Chagaluka, G., Adlar, F.R., Offor, U.T., Broadbent, C., Chaytor, L., Whitehead, A., Hall, A., O'Connor, H., Van Noorden, S., Lampert, I., Bailey, S., Molyneux, E., Bacon, C.M., Bomken, S. and Rand, V. (2019) 'Sporadic and endemic Burkitt lymphoma have frequent FOXO1 mutations but distinct hotspots in the AKT recognition motif', *Blood Adv*, 3(14), pp. 2118-2127.

Zhu, K.Y., Song, K.W., Connors, J.M., Leitch, H., Barnett, M.J., Ramadan, K., Slack, G.W., Abou Mourad, Y., Forrest, D.L., Hogge, D.E., Nantel, S.H., Narayanan, S., Nevill, T.J., Power, M.M., Sanford, D.S., Sutherland, H.J., Tucker, T., Toze, C.L., Sehn, L.H., Broady, R. and Gerrie, A.S. (2018) 'Excellent real-world outcomes of adults with Burkitt lymphoma treated with CODOX-M/IVAC plus or minus rituximab', *Br J Haematol*, 181(6), pp. 782-790.